A Regression Discontinuity Test of Strategic Voting and Duverger's Law[*]

Thomas Fujiwara

Department of Economics, Princeton University, USA;
fujiwara@princeton.edu.

ABSTRACT

This paper uses exogenous variation in electoral rules to test the predictions of strategic voting models and the causal validity of Duverger's Law. Exploiting a regression discontinuity design in the assignment of single-ballot and dual-ballot (runoff) plurality systems in Brazilian mayoral races, the results indicate that single-ballot plurality rule causes voters to desert third placed candidates and vote for the top two vote getters. The effects are stronger in close elections and cannot be explained by differences in the number of candidates, as well as their party affiliation and observable characteristics.

Political scientists and economists have long been interested in the question of whether citizens vote sincerely or strategically. How voting decisions take place is not only fundamental to the understanding of the democratic process, but also has important implications for the formulation of

theory. Virtually any formal model with voting for three or more candidates requires the *assumption* that voters act either sincerely or strategically, and this choice usually has important implications for the model's results and conclusions.[1]

The best-known prediction regarding strategic voting in a multi-candidate setting is Duverger's Law, named after Duverger's (1954) prediction that "*simple-majority single-ballot* [plurality or first-past-the-post rule] *favors the two party system*" whereas "*simple majority with a second ballot* [dual-ballot or runoff] *or proportional representation favors multipartyism.*"[2]

In this paper, I empirically test Duverger's Law and address the validity of its *causal* statement by exploiting a regression discontinuity design in the assignment of electoral rules in Brazilian municipal elections. I also investigate the mechanisms that drive the association between plurality and two-party dominance, and provide evidence that suggests that strategic voting is the most likely driving force behind the results.

Duverger's rationale was that single-ballot plurality rule[3] creates an incentive for voters to engage in a particular pattern of strategic voting, which can be described by an example. A citizen believes that candidates 1 and 2 have the highest probability of winning an election (and that a tie between 1 and 2 is more likely than a tie between any other two candidates). His preferred choice, however, is candidate 3. To maximize his chances of being a pivotal voter, he strategically chooses to vote for his preferred choice between 1 and 2. As all voters go through a similar logic, candidate 3 is deserted by her supporters, which all vote for candidates 1 and 2.[4]

Duverger also argued that this rallying behind the two top candidates would not occur under dual-ballot plurality (also known as the runoff or two-round electoral rule), a system where voters may go to the ballot box twice. First, an election is held and if a candidate obtains more than 50% of the votes, she is elected. If not, then a second round of voting is held where

---

[1] A compelling example is the citizen-candidate models of Osborne and Slivinski (1996) and Besley and Coate (1997). The structure of the two independently developed models is similar; however, but the latter assumes strategic voting whereas the former assumes sincere voting, which results in different equilibrium policies.

[2] Riker (1982) discusses the history of Duverger's Law and its status as "*a true sociological law.*"

[3] Single-ballot plurality is also referred as plurality rule or first-past-the-post. It is the system where the candidate with the most votes is elected, such as the one used for the U.S. House of Representatives and the U.K.'s House of Commons.

[4] In this paper, all voters are male and all candidates female.

only the two most voted candidates in the first round face off.[5] In the first round of a dual-ballot system, a strategic voter would still find worthwhile voting for a candidate that he expects to finish in third place (as he could be pivotal in pushing her to the second round).

Duverger's argument is formalized by multiple game-theoretic models of voting and tested in a number of papers that compares electoral results under single- and dual-ballot rules. Results are mixed as Wright and Riker (1989) and Golder (2006) finds support for it, while Shugart and Taagepera (1994) and Engstrom and Engstrom (2008) do not.

The endogeneity of electoral rules is an obstacle to the causal interpretation of the results in these papers. Regions with different electoral rules are likely to also differ in other (observed or unobserved) characteristics that also affect electoral results. Moreover, there is also the possibility of *"a causality following in the reverse direction, from the number of parties towards electoral rules"* (Taagepera, 2003).[6]

This paper provides a cleaner test of Duverger's Law that does not suffer from these issues by focusing on a natural experiment where electoral rules are exogenously assigned. The Brazilian Constitution mandates that municipalities with less than 200,000 registered voters use single-ballot plurality rule to elect their mayors, while those with an electorate size above such threshold should use dual-ballots. This regression discontinuity design generates assignment of electoral rules that is as good as random and allows causal inference of its effects. Intuitively, municipalities just below the threshold should be, on average, similar in all observed and unobserved characteristics to those just above it, so that any difference in outcomes between these two groups must be caused by the different electoral rules.

Results based on data on the outcomes of the universe of Brazilian mayoral elections in the 1996–2008 period show that, as predicted by Duverger's Law, a change from single-ballot to dual-ballot decreases voting for the top two vote getters and raises it for the third and lower placed candidates.[7]

---

[5]   Dual-ballot is the most used electoral system for presidential elections in the world (Golder, 2006). It is common in primaries in the Southern United States and several large American cities, as well as regional elections in France, Italy, and Switzerland. In some cases, the threshold for first-round victory differs from 50%.

[6]   The argument is that societies with a predisposition to the existence of multiple parties are likely select an electoral system that is more suited to accommodate them.

[7]   Throughout the paper, voting under dual-ballot refers only to voting in the first round.

Moreover, this effect is stronger in closely contested races where the incentives to vote strategically are larger.

While the above results validate the empirical content of Duverger's Law, it leaves open the possibility that the observed effect of electoral rules on voting is driven by channels that are unrelated to strategic voter behavior. Even with a sincere electorate, the results could be observed if different electoral rules generate systematic differences in candidates' characteristics and behavior.

To explore the plausibility of these competing mechanisms, I provide a series of results that help rule them out and strengthen the case that strategic voting is the driving force behind the results. I find that the exogenous change from single- to dual-ballot does not affect several observed characteristics of mayoral candidates (party affiliation, education, occupation). To address the issue of (unobserved) candidate behavior, I exploit that mayoral elections occur on the same day as municipal legislature elections, and that the electoral rule for the latter (proportional representation) is the same in all municipalities. If mayoral candidate behavior is the driving force behind the results, then such behavior would likely also affect legislature elections through a coattail effect.[8] I find no evidence for this type of spill-over effect, as election results in legislative elections are not systematically affected by the exogenous change in mayoral electoral rule.

A similar regression discontinuity design in the assignment of single- and dual-ballot rules in Italian municipalities is also exploited by Bordignon *et al.* (2010), who find that dual-ballot leads to a larger number of candidates and smaller policy variability than single-ballot, conforming to the predictions of a model of party formation. Although this paper focus on a different issue (voter behavior), I discuss some of the similarities in the results.[9]

This paper also communicates with the literature that measures the extent of strategic voting, which includes small-scale laboratory experiments (surveyed in Rietz, 2008) and surveys analyses that directly ask respondents

---

[8]   For example, if a third placed candidate has increased voting under dual-ballots solely because she campaigns more intensively under such rule, one would expect the legislature candidates from her party to also benefit from this additional campaigning.

[9]   Short after the preparation of the original draft of this manuscript, I became aware of an independently developed paper (Chamon *et al.,* 2009) that explores the same regression discontinuity design, but focuses mainly on the effect of electoral rules on fiscal spending. Another independently developed paper (Gonçalves *et al.*, 2008) explores if dual-ballots increase the number of candidates that enter Brazilian mayoral races, using a difference-in-differences approach instead of the regression discontinuity design.

about their preferences and votes (Alvarez and Nagler, 2000).[10] While the former approach must deal with the difficulties in how to elicit preferences and the measurement issues related to survey questions about previous voting behavior (Wright, 1990, 1992; Mullainathan and Washington, 2009), the latter approach leaves open the question whether strategic voting occurs in elections with electorates thousands of times larger than the ones used in experiments.

Theoretical Framework

While the term "Duverger's Law" is open to multiple interpretations, for the sake of clarity and specificity this paper associates the term with a particular hypothesis that can be empirically tested: *compared single-ballot plurality, dual-ballot reduces the vote share of the two most voted candidates.* Note that this is a causal statement that does not refer only to an empirical correlation between electoral rules and voting.

The theoretical underpinning of the hypothesis above is provided by several papers that analyze game-theoretic models of strategic voting under single-ballot rule, such as Palfrey (1989), Myerson and Weber (1993), Cox (1994), Myerson (2002), and Myatt (2007). The case of dual-ballot rule is studied by Cox (1997), Martinelli (2002), and Bouton (2011).[11]

These models predict that, under single-ballot, there exists an equilibrium where only two candidates receive all the votes, with the remaining candidates being strategically abandoned by their followers, due to a mechanism that can be intuitively described by the example on the introduction. Under dual-ballots, there exists an equilibrium where only three candidates receive a strictly positive amount of votes in the first round. Hence, the empirical hypothesis being tested in this paper is borne out by this theoretical literature.

There are, however, some caveats to be made in the mapping from theoretical analysis to empirical formulation. First, a common feature of these models is the presence of multiple equilibria. Under single-ballot electoral rules, there can also exist an additional equilibrium where the third placed candidate also receives a positive amount of votes.[12] Hence, these models do

---

[10] Additionally, Kawai and Watanabe (2010) estimate a structural model of voting with aggregate vote shares. Degan and Merlo (2009) analyze a different kind of strategic voting (split tickets).

[11] Cox (1997) and Myerson (1999) survey this literature.

[12] Moreover, Bouton (2011) presents a case of an equilibrium where only two candidates receive a positive amount of votes under dual-ballot rules.

not have clear-cut predictions that can be directly tested without making specific assumptions on equilibrium selection.

Second, some implications of these models are too simplified or stark to be taken to data directly. While the models feature a complete abandonment of the third and lower placed candidate under single-ballot (i.e., zero votes in equilibrium), one would not expect to observe a candidate that no one votes for in an actual election. Hence, the hypothesis being tested deals with some amount of strategic abandonment of the third placed candidate, and not a complete desertion by her supporters. Myatt (2007) is a notable exception that provides a model with a unique equilibrium where the third placed candidate does suffer from strategic desertion but still receives a positive amount of votes.[13] However, the model is characterized only under single-ballots (and for the case of three candidates).

Another explanation for the non-absolute abandonment of the third placed candidates is that a combination of sincere and strategic voters co-exist, with the former type guaranteeing that all candidates receive a positive amount of votes. Although this is a reasonable possibility, the theoretical analyses cited above deal only with cases where all voters are strategic. Moreover, it is also possible that other types of voter behavior are also present. One case would be bandwagon effects, where citizens have a preference for voting for the winner (Simon, 1954). The presence of bandwagon voters would likely reinforce the Duverger's Law effect described above (i.e., under single-ballot plurality, the candidate expected to have the most votes benefitting from the strategic abandonment of third placed candidate would add impetus for bandwagon voters who wish to vote for the winner).[14]

Empirical Strategy

Brazil is constituted by more than 5000 municipalities, which are the smallest level of government in the country (similar to an American town or city).

---

[13] The model uses the global games approach to obtain a unique equilibrium in coordination games.

[14] Note, however, that a population of only sincere and/or bandwagon voters would not clearly lead to the hypothesis that is tested in this paper: that dual-ballots reduce the vote share of the two most voted candidates. It is not clear that, absent strategic voting, dual-ballots would lower the probability that the candidates expected to finish first are the winner, while raising it for the one expected to finish third. In other words, Duvergerian strategic voting predicts a specific pattern for the effect of dual-ballots across candidates finishing the race in different positions that is distinct from that of bandwagon voters.

Each municipality has a single mayor (*Prefeito*) and a municipal legislature (*Câmara de Vereadores*), which are elected every four years. Municipal elections are regulated by federal legislation, and all municipalities have the same election and inauguration dates. Municipalities are not divided into districts, so that elections are at large.

Brazilian legislation requires all citizens aged 18 or older to register to vote in their municipality of residence. Moreover, the Constitution states that mayoral elections should be run under the single-ballot plurality rule system (SB, henceforth) in municipalities with less than 200,000 voters, while municipalities with 200,000 voters or more must have their elections under dual-ballot plurality rule (DB, henceforth).

This threshold-based rule creates a standard regression discontinuity design. Under mild assumptions, it generates quasi-random assignment because municipalities just below and just above the threshold should be, on average, *ex ante* similar to each other in every possible aspect. In other words, the reason that they are on a particular side of the threshold is due to random uncontrollable events that should not be related to the outcome of interest. This argument is formalized by Lee (2008). Other than voting rule, any observed or unobserved variable that could affect voting should be the same for all municipalities that are sufficiently close to the threshold. This guarantees that any difference in outcomes between these two groups is a causal consequence of the different electoral rules.

For this to hold, it is important that the 200,000-voter threshold is somewhat arbitrary and not used to assign anything else to municipalities. To the best of my knowledge, this is the case. Although some other regulations of municipal governments depend on its population (which is different from its number of voters), none of them has threshold close to 200,000 voters.

The cutoff is established by the Brazilian Constitution (ratified in 1988). The likely reason for the rule was that, although dual-ballot was deemed superior to SB by the Constituent Assembly,[15] the cost of a possible second round of elections in the universe of municipalities was prohibitive. Moreover, even if the cutoff was set aiming to keep a particular group of municipalities under a particular electoral rule, by 1996 (when the first election in the sample was held) the different rates of population and registration growth between municipalities would likely have dissipated this effect.

---

[15] The constitution dictates that all state governors and the president must be elected by DB.

This paper uses data provided by the federal elections authority (*Tribunal Superior Eleitoral*) on election results, candidate information, and electorate characteristics (e.g., turnout, registration) for *all* Brazilian municipalities in the 1996, 2000, 2004, and 2008 elections.[16] The unit of observation is a municipality-election and there are over 20,000 observations in the sample, although most estimates use substantially smaller samples that include only observations close to the 200,000-voter threshold. A table with descriptive statistics can be found in Appendix A.

Analysis of the data shows that there was full compliance with the assignment rule, as no municipality with less than 200,000 voters had a second round of votes and all municipalities with more than 200,000 voters where no candidate obtained more than 50% of the votes in the first round of election had a second round of election. Hence, the regression discontinuity design is sharp (i.e., the probability of treatment changes from zero to one at the threshold).

Another threat to validity would occur if a change from SB to DB affected turnout. If different groups of voters attend the polls under the different electoral rules, then the research design may not be able to successfully compare similar electorates under different electoral rules. Fortunately (for the paper's research design), Brazilian law makes registration and voting compulsory for all citizens aged 18–70.[17] Failing to register or vote in a previous election renders a citizen ineligible to several public provided services until a fine is paid. Moreover, elections are held on a Sunday and voters are allocated to polls close to their residence in order to foster turnout. Although these features do not guarantee a turnout close to 100% in the elections,[18] it makes the issues related to election outcomes a second-order issue in the decision to vote or not and hence the observed difference in turnout under SB and DB is virtually zero.

Another issue is the possibility of strategic manipulation of the forcing variable. If, for any reason, some agent (such as a party or the government) had a preference for SB or DB, it could try to manipulate the registration of voters in order to fall on the preferred side of the threshold. This kind of behavior would likely invalidate the analysis, because some amount of

---

[16] Detailed data for previous elections are not available.
[17] Voting is voluntary for citizens aged 16–17 or 70+ and to those who are officially illiterate.
[18] Figure 2 and Table 1 show that turnout is in the order of 85% of registered voters. This occurs because citizens who are not in their city of residence on election day can be waived from the punishment by attending a poll in any other municipality and submitting a waiver form.

self-selection would occur between SB and DB rules. However, if strategic registration does indeed take place, it would likely be reflected in a discontinuity in registration rates or in the number of cities that are above or below the threshold. Both of these issues can be tested (and rejected) in the data.

Finally, some of the implications of a regression discontinuity design's quasi-random assignment can be tested. In randomized controlled experiments, it is usual to check for the possible failure to randomize by comparing predetermined variables on treatment and control groups. Similarly, in the regression discontinuity context it is possible to test if the treatment effect in outcomes that were determined before the assignment of electoral rules is zero.

Estimation Framework

Let v be the number of registered voters in a municipality. The treatment effect of a change from single-ballot to dual-ballot on outcome y is given by

$$TE = \lim_{v \downarrow 200,000} \mathrm{E}[y|v] - \lim_{v \uparrow 200,000} \mathrm{E}[y|v].$$

Under the assumption that the conditional expectation of y on v is continuous, the first term on the right-hand side converges to the expected outcome of a municipality with 200,000 voters and DB, while the second term converges to the expected outcome of a municipality with 200,000 voters and SB. Hence, TE identifies the treatment effect of changing from SB to DB for a municipality of 200,000 voters, as long as the distribution of treatment effects is continuous at the threshold.

The estimation method used here closely follows the guidelines in Imbens and Lemieux (2008), which in turn rely on the results provided by Hahn *et al.* (2001). The reader is referred to these papers, as only a brief overview is provided here. The limits on the right-hand side are estimated non-parametrically by local polynomial regression. This consists of the estimation of a linear (or quadratic) regression[19] of y on v with only data that satisfies $v \in [200,000 - h; 200,000]$. The predicted value at $v = 200,000$ is thus an estimate of the limit of y as $v \uparrow 200,000$. Similarly, a regression with only data that satisfies $v \in [200,000; 200,000 + h]$ is used to estimate the limit of y when $v \downarrow 200,000$. The difference between these two estimated

---

[19] Notice that the regression is unweighted (i.e., rectangular kernel).

limits is the treatment effect. It is important to note the non-parametric nature of the estimation: although linear (or quadratic) regressions are used, the consistency of the results holds for any arbitrary and unknown shape of the relationship between y and v. The limit approaching one side of the threshold is estimated with only data on that particular side.

The local linear regression estimate can be implemented by OLS estimation of the following single equation using only observations that satisfy $v \in (200,000 - h; 200,000 + h)$.

$$y = \alpha + \beta(v - 200,000) + \gamma \cdot 1\{v > 200,000\}$$
$$+ \delta(v - 200,000) \cdot 1\{v > 200,000\} + u,$$

where $1\{v > 200,000\}$ is a dummy variable that takes value one if, and only if, the election is carried under DB, u is the error term, and the parameters to be estimated are denoted in Greek letters. The estimate of γ is the treatment effect and its (heteroskedasticity and cluster-robust) standard error can be obtained in a straightforward manner. The estimation with a quadratic specification just adds two more variables: the square value of v and its interaction with $1\{v > 200,000\}$. To control for election-year specific effects, a set of dummies that indicate the year in which the election took place is included in all estimations in the paper.

A key decision is h, the kernel bandwidth. Higher values generate more precision but create larger bias. To show the robustness of the results to different choices of h, this paper presents the results for three different levels: 25,000, 50,000, and 75,000 voters. Note that these are relatively small and hence try to reinforce the local intuition of regression discontinuity designs: although there are more than 20,000 observations (municipal elections) in the data, less than 300 are used to obtain all of the estimates.

Note that, given the size distribution of Brazilian municipalities, as the bandwidth increases, the number of (smaller) municipalities that are included at the extreme of the left interval increases rapidly (see Figure 3). Hence, estimates with large bandwidths are likely to put too much weight on the fit of the relationship away from the neighborhood of the 200,000 threshold.

Main Results

I start with the main result that a change from SB to DB increases the vote share of the third and lower placed candidates. The following sections
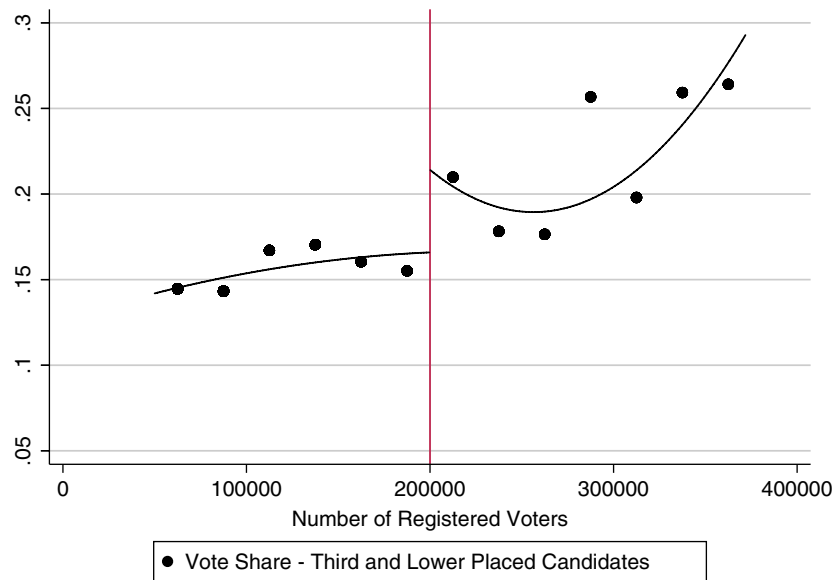
Figure 1. Vote share of third and lower placed candidates — local averages and parametric fit.

provide the evidence in favor of the quasi-random nature of the assignment of electoral rules and other robustness checks.

Figure 1 presents the share of votes that were received by the third and lower placed candidates, against the forcing variable (number of registered voters). Each point in the figure reflects the average outcome for a bin of municipalities that fall within a 25,000-wide interval of the forcing variable. For example, the first circle to the right of the vertical line that indicates the 200,000-voter threshold equals the average vote share of third and lower placed candidates in municipalities with $v \in [200,000; 225,000]$. To facilitate visualization, a quadratic model is fitted at each side of the 200,000 threshold, so that the point where the lines are not connected is where the discontinuity in outcomes, if existent, is expected to be visible.[20]

While the relationship is smooth to the left of the 200,000-voter line, there is a jump right after cutoff value. The fitted curves indicate that the vote share for the third or lower placed candidates in the election increases from

---

[20] In some graphics, a quadratic relationship is fitted, whereas in others a linear one is used. The decision on which one to use is made by one specification against the other.

Table 1. Treatment effects on electoral outcomes.

Specification/ bandwidth	Single-ballot mean	Linear 50,000	Linear 25,000	Linear 75,000	Quad. 50,000	Quad. 75,000
Dependent variable		(1)	(2)	(3)	(4)	(5)
Vote share — 3rd and	0.155	0.088	0.093	0.069	0.104	0.113
lower placed candidates		(0.040)	(0.056)	(0.033)	(0.058)	(0.046)
Vote Share — 4th and	0.041	0.043	0.046	0.036	0.057	0.055
lower placed candidates		(0.024)	(0.030)	(0.021)	(0.031)	(0.028)
Vote Share — 5th and	0.012	0.015	0.017	0.015	0.022	0.021
lower placed candidates		(0.010)	(0.012)	(0.009)	(0.012)	(0.011)
Registration rate	0.638	0.011	0.016	0.021	0.031	0.014
		(0.019)	(0.030)	(0.016)	(0.029)	(0.024)
Turnout rate	0.851	0.003	−0.004	0.002	−0.003	−0.002
		(0.007)	(0.011)	(0.007)	(0.01)	(0.009)
Observations	—	175	81	282	175	282

Robust standard errors clustered at the municipality level in parenthesis. Each figure in the table is from a separate local linear/quadratic regression with the specified bandwidth. The level of observation is a municipal election. The estimated treatment effect is of a change from SB to DB. All estimates include year effects. Details on the dependent variables are presented in the text.

about 15 p.p. to 23 p.p. as there is a change from SB to DB. In other words, DB increases voting for the third (and lower) candidates by roughly 50%.[21]

The formal estimate counterparts of the depicted jump are provided in the first row of Table 1. Columns (1)–(3) present the results for different bandwidths with the local linear regression, whereas columns (4) and (5) probe the robustness of the result with a quadratic specification. Throughout the paper, the estimate presented is the treatment effect of a change from SB to DB. In program evaluation jargon, DB is the treatment and SB is the control. To help evaluate the magnitude of the effects, the *single-ballot mean* — the average for municipalities within a 25,000-voter interval below the 200,000-voter threshold — is presented. All standard errors are clustered at the municipality level and hence are robust to serial correlation of unknown form.

---

[21] Figure 1 also shows that the relationship between vote share and the number of voters is noisier at the right side of the cutoff. This occurs as, given the size distribution of municipalities (Figure 3), there are progressively less observations in each bin as the number of voters gets larger.

The 0.088 figure presented in the first row of column (1) hence indicates that a change from SB to DB increases the vote share of third and lower placed candidates by 8.8 percentage points. This effect is significant at the 5% level and implies a large positive effect (a 56% increase from the 15.5 p.p. single-ballot mean). This is consistent with more than half the voters who would vote for the third and lower placed candidate under DB strategically deserting her and voting for the top two candidates under SB. Columns (2)–(5) show that the numerical estimate and its statistical significance are not meaningfully affected by different choices of bandwidths or the use of a quadratic specification. Appendix B shows that the estimates above are also robust to the inclusion of several different covariates.

The second row repeats the same exercise for the vote share of the fourth and lower placed candidates. The estimates are usually less than half of its counterpart in the first row (and usually significant at the 10% level). The third row of Table 1 addresses the effects on the vote share of the fifth and lower placed candidates in similar fashion, as the estimated effects are numerically close to (and statistically indistinct from) zero.

This set of results indicates that virtually all the votes that the top two candidates lose when changing from SB to DB are gained by the third and fourth placed candidates, with their majority going to the third placed candidate. Note that the difference between estimates in the first and second row equal the treatment effect on the vote share of the third placed candidate, while the difference between the second and third row is equal to the effect for the fourth placed candidate.[22]

In order to assess the threats to validity, Figure 2 repeats the exercise of Figure 1 for the turnout rate (total turnout divided by the number of registered voters) and the registration rate (ratio of registered voters to the total population in the municipality). The relationship between these variables and the number of voters is smooth and does not present a jump at the threshold. Hence, the increase in votes for third and lower placed candidates is not driven by differences in turnout in SB and DB municipalities, just as

---

[22] Although the estimates of the effect on fourth and lower placed candidate vote share are not very precise, they do provide some evidence that the fourth placed candidate also benefits from a change from SB to DB. Such possibility is supported by the theoretical model of Cox (1997) where under DB elections there is both a type of equilibria where fourth candidate receives zero votes and a type where he receives the same amount as the third candidate. The intuition is that under the beliefs and expectations of a tie, voters do not know on which candidate to strategically coordinate on, making the expectation of a tie self-fulfilling. An analogous result for the case of the second and third placed candidates under SB also holds.
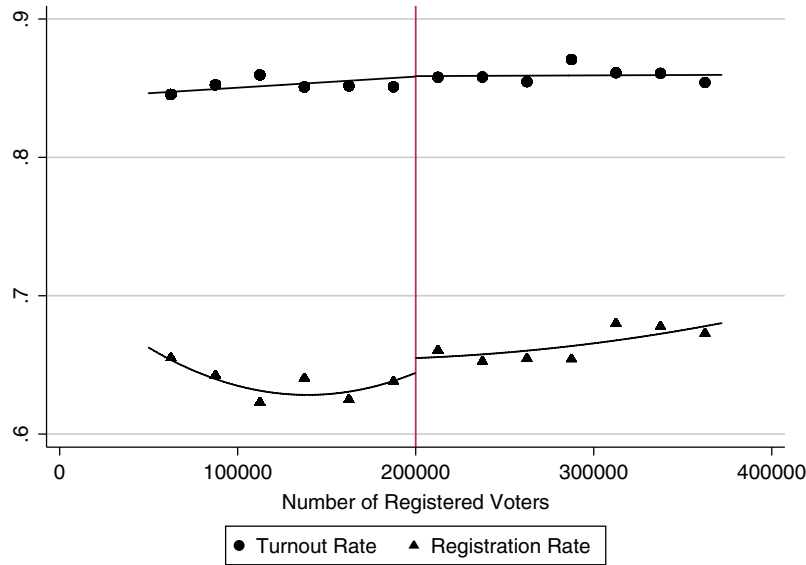
Figure 2. Turnout and registration — local averages and parametric fit.

there is no evidence that strategic manipulation of the number of registered voters has taken place. The formal counterpart is provided in the fourth and fifth rows of Table 1, which show that the estimated treatment effects on turnout and registration are numerically small and statistically insignificant.

To further probe the possibility of strategic manipulation, Figure 3 implements an exercise suggested by McCrary (2008) and plots the number of observations contained in each bin of the previous figures. If strategic manipulation has taken place, it would likely reflect in a jump close to the threshold. If, for example, governments in municipalities just below the threshold tried to deter registration in order to avoid switching to DB in the near future, then the number of municipalities just below the threshold would probably be unusually large compared with the number of municipalities just above. As Figure 3 shows, such jump is not observed and hence no evidence of strategic manipulation is found.

Tests for Quasi-Random Assignment

The intuition of the identification strategy is that SB and DB systems in elections close to the threshold are assigned quasi-randomly, so that
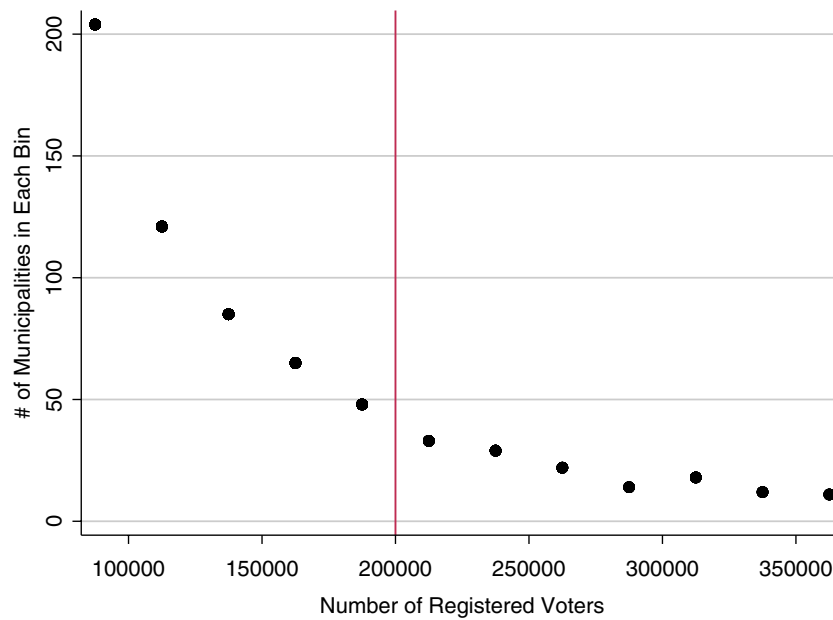
Figure 3. Distribution of electorate size.

municipalities just below and just above the threshold are similar in all observed and unobserved predetermined characteristics.

Although there is good reason to believe that this is indeed the case, Table 2 provides some evidence that this intuition holds. It does so by checking if the values of baseline characteristics that should not be affected by electoral rules are similar on each side of the 200,000-voter threshold. In other words, I estimate treatment effects where they are expected to be zero, an exercise that is analogous to the common practice of testing for randomization in controlled experiments by comparing averages of baseline variables in the treatment and control group.

Table 2 presents the estimated treatment effects for a host of geographic and economic variables: the municipalities' longitude and latitude (measured in degrees), *per capita* monthly income (in 2000 *reais*), income inequality (measured by the Gini index), education (average years of schooling in the population aged 25 or older) and the population share living in a rural area. The source of all these variables is the Brazilian statistical agency (*Instituto Brasileiro de Geografia e Estatística*).

Table 2. Tests of quasi-random assignment.

Specification/ bandwidth	Single-ballot mean	Linear 50,000	Linear 25,000	Linear 75,000	Quad. 50,000	Quad. 75,000
dependent variable		(1)	(2)	(3)	(4)	(5)
Longitude	47.203	2.057	4.529	1.048	3.543	2.416
(in degrees)		(1.441)	(2.515)	(1.181)	(2.258)	(1.964)
Latitude	−19.540	−2.379	−4.624	−1.997	−4.42	−2.851
(in degrees)		(1.785)	(3.005)	(1.416)	(2.744)	(2.261)
Per capita	316.401	9.766	19.035	31.126	4.986	−8.971
Income (R$)		(24.769)	(41.06)	(24.391)	(36.913)	(34.345)
Gini index for	0.554	−0.009	−0.004	−0.010	0.001	−0.006
Per capita income		(0.012)	(0.019)	(0.011)	(0.018)	(0.015)
Years of	6.323	0.112	0.278	0.236	0.219	0.044
schooling		(0.217)	(0.355)	(0.189)	(0.295)	(0.285)
Pop. Share	0.048	−0.008	−0.008	−0.016	−0.020	−0.007
in rural areas		(0.013)	(0.02)	(0.013)	(0.017)	(0.014)
Observations	—	175	81	281	175	281

Robust standard errors clustered at the municipality level in parenthesis. Each figure in the table is from a separate local linear/quadratic regression with the specified bandwidth. The level of observation is a municipal election. The estimated treatment effect is of a change from SB to DB. All estimates include year effects. Details on the dependent variables are presented in the text.

These variables were observed only for the census years of 1991 and 2000. I assign the value from a previous census to each municipality-election observation (i.e., data from the 1991 Census is assigned to the 1996 elections and data from the 2000 Census is assigned to the 2000, 2004, and 2008 elections). The estimated treatment effects are numerically small and statistically insignificant, independently of the bandwidth or specification used in the estimation. This set of results indicate that municipalities just below and just above the cutoff are similar in several dimensions, supporting the quasi-random interpretation of the effects in Table 1.

Effects in Contested and Uncontested Elections

In elections in which one candidate is expected to win for sure, there is presumably less reason to act strategically. Hence, in elections that are perceived as practically uncontested, the effect of a change from SB to DB on the vote share of the third and lower placed candidates should be smaller.

In a formal model, this phenomenon can be represented by the equilibria where expectations are such that the probability of a tie between the first-placed candidate and any other candidate is exactly zero and hence the (expected) probability of a voter being pivotal is zero no matter who he votes for.[23]

To capture this intuition, the sample is split into a *contested* and *uncontested* elections subsamples. The former are those where the winner obtains less than 50% of the votes (in the SB election or on the first round of the DB election), whereas the latter includes those where the winner obtained a majority.

The 50% mark captures two important features. First, in the uncontested elections even if all voters that did not vote for the winner coordinated perfectly and voted for some other candidate, the results of the election would remain unchanged. Second, the *uncontested* elections are those where there is no second round under DB.

However, the vote shares of the first-placed candidate and the dependent variable of interest (vote share of third and lower placed candidates) are mechanically correlated, raising the econometric concerns related to sample selection biases. To sidestep these issues, I use the vote shares predicted by previous elections results (i.e., lagged variables) to split the sample.

I do so first by estimating a logit regression of the indicator for contested status against a lagged contested status, (also lagged) vote share of the first placed candidate, and a set of year dummies. I then use the fitted probabilities from this model to assign *predicted to be contested* and *predicted to be uncontested* status for the elections, with the latter (former) being when the fitted probability is above (below) 50%. This procedure separates the sample based only on variation from lagged variables, and hence avoids sample selection based on a variable mechanically correlated to the dependent variable.

The results are presented in Table 3. Panel A repeats the estimation reported on the first row of Table 1 with only the sample of elections predicted to be contested.[24] The estimates are large and usually statistically significant. Panel B provides the counterpart from the sample with elections predicted to be uncontested. The estimates are numerically close to zero and statistically indistinct from it. These results imply that the effect of DB

---

[23] Note, however, that in the only paper that provides unique equilibria for the SB case (Myatt, 2007), this prediction does not hold.

[24] Note that, because of the use of lagged variables, the first year of the sample (1996) was dropped. Hence, the total sample sizes are smaller than in Tables 1 and 2.

Table 3. Treatment effects in contested and uncontested elections.

Specification/ bandwidth	SB mean	Linear 50,000	Linear 25,000	Linear 75,000	Quad. 50,000	Quad. 75,000
		(1)	(2)	(3)	(4)	(5)
Panel A: Elections predicted to be contested						
Vote share — 3rd and lower placed candidates	0.148	0.157 (0.076)	0.145 (0.107)	0.144 (0.061)	0.145 (0.081)	0.177 (0.083)
Observations	—	64	25	109	64	109
Panel B: Elections predicted to be uncontested						
Vote share — 3rd and lower placed candidates	0.138	0.015 (0.049)	0.001 (0.075)	0.011 (0.039)	0.003 (0.075)	0.032 (0.057)
Observations	—	80	40	123	80	123

Robust standard errors clustered at the municipality level in parenthesis. Each figure in the table is from a separate local linear/quadratic regression with the specified bandwidth. The level of observation is a municipal election. All estimates include year effects. Details on the dependent variables are presented in the text.

on the vote share of the third placed candidate is almost entirely driven by the elections predicted to be contested, supporting the intuition that voters are less likely to act strategically in elections where doing so is less likely to matter.

Competing Mechanisms

The set of results discussed in the main results section indicates that a change from SB to DB increases the vote share of the third and lower placed candidates. However, it leaves open the possibility that the observed effect of electoral rules on voting is driven by channels that are unrelated to strategic voter behavior. Even with a sincere electorate, the results could be observed if different electoral rules generate systematic differences on other factors:

- The number of candidates. For example, it could be that DB races are less likely to be two candidate races, creating a mechanic association between electoral rules and vote shares.

- Party affiliation of the contesting candidates, as it would be possible that a competitive party that tends to finish in third place is more likely to enter in DB elections, for example.
- The quality of candidates. If third placed candidates that run under DB are more attractive to voters, its effect on vote share could be observed even if all voters are sincere.
- Candidate behavior, as it could be possible that third placed candidates campaign more intensively under DB.

The following sections address these issues separately, providing evidence suggesting that each of these possibilities can be ruled out, leaving strategic voting as the most likely explanation. It must be noted that providing *direct* evidence to rule out differential *unobserved* candidate quality and behavior under SB and DB is (by the definition of unobserved) not possible, so that the third and fourth mechanisms cannot be ruled out with certainty. However, an indirect test for their roles is provided, and while the overall combination of results in this paper can straightforwardly be explained by strategic voting, it would require a more convoluted argument based on one of the factors listed above.

Number of Candidates

The *vote share of ith and lower placed candidates* variables analyzed in Table 1 is defined in such way that they are equal to zero in election with less than i candidates (e.g., in a three candidate race, the vote share of the fourth and lower placed candidates is zero). This raises the possibility that DB increases the vote share of the ith placed candidate by the mechanical reason that DB elections are more likely to have an ith candidate.

In the samples used to estimate the treatment effects in the previous section, all elections have at least two candidates. Some races, however, have less than three candidates; hence, it is possible to estimate the effect of a change from SB to DB on the probability of three or more candidates running in the election. This is done by the addition of a dummy indicator taking a value of one if the election has at least three candidate in as the dependent variable in a regression similar to the ones performed in Tables 1 and 2.

The first row of Table 4 provides such estimates. It must be noted that the single-ballot mean is 96%, so that almost all races under SB close to the threshold have three or more candidates. The estimated treatment effects

Table 4. Treatment effects on number of candidates.

Specification/ bandwidth	Single-ballot mean	Linear 50,000	Linear 25,000	Linear 75,000	Quad. 50,000	Quad. 75,000
Dependent variable		(1)	(2)	(3)	(4)	(5)
Indicator for candidates ≥ 3	0.958	0.038 (0.037)	0.027 (0.056)	0.034 (0.026)	−0.006 (0.052)	0.035 (0.039)
Indicator for candidates ≥ 4	0.833	0.118 (0.115)	0.125 (0.165)	0.043 (0.087)	0.099 (0.187)	0.116 (0.140)
Indicator for candidates ≥ 5	0.479	0.269 (0.144)	0.316 (0.201)	0.235 (0.124)	0.334 (0.202)	0.286 (0.170)
Number of candidates	4.792	0.706 (0.463)	0.984 (0.624)	0.738 (0.402)	1.017 (0.655)	0.679 (0.552)
Observations	—	175	81	282	175	282

Robust standard errors clustered at the municipality level in parenthesis. Each figure in the table is from a separate local linear/quadratic regression with the specified bandwidth. The level of observation is a municipal election. All estimates include year effects. Details on the dependent variables are presented in the text.

are numerically and statistically close to zero, which implies that a change to DB does not affect the probability of a third candidate entering the race. This result implies that the number of candidates cannot explain the results on the previous section. In other words, DB increases the vote share of third placed candidates but does not increase the probability that a third candidate enters the race.

To further characterize the impact of DB on the number of candidates, the second and third rows of Table 4, respectively, present the estimated treatment effect on the probability of the number of candidates being four or larger. While the effects on the former are relatively small and statistically insignificant, the effects on the latter are larger and usually significant at the 5% and 10% level, depending on the bandwidth and specification. This set of results indicates that DB raises the number of candidates through an increase in the probability that a race has five or more contestants, and not by the addition of a third or fourth candidate.

As the results discussed in the previous section show that DB only affects the vote shares of the third and (to a lesser extent) fourth placed candidates,

it is possible to rule out the possibility that these effects on vote share are mechanically driven by the entry of a different number of candidates under SB and DB races.

These results are also of interest because of theoretical analyses which propose that DB increases the number of candidates that enter the race compared with SB (Osborne and Slivinski, 1996; Bordignon *et al.*, 2010). Moreover, the latter study[25] finds that there is indeed a larger number of candidates under DB than SB exploiting a similar regression discontinuity design in Italian municipalities. To facilitate comparison to the results in Bordignon *et al.* (2010), the last row of Table 4 presents the treatment effect on the total number of candidates. The average SB race close to the cutoff has 4.8 candidates, and DB seems to add a 0.7–1.0 candidate to the race, although this effect is never significant at the 10% level.[26] Figure 4 plots the number of candidates against the number of registered voters, where a relatively small jump at the threshold and an upward trend are observed.[27]

These results are of a slightly smaller magnitude than those in Bordignon *et al.* (2010). In the Italian case the threshold involves substantially smaller municipalities (15,000 residents) and a smaller number of candidates under SB (about 3.7 candidates close to the cutoff), with their estimates in the range of 1.0–1.5 additional candidate from a change DB. Given the relative imprecision of the estimates in Table 4, this and Bordignon *et al.*'s (2010) papers point out similar conclusions with regards to the effect of DB on the number of candidates. Aside the issue of precision, a likely explanation for the larger effect the Italian case is that the smaller number of candidates makes the strategic consideration to enter a mayoral race more dependent on the electoral rule. However, city size and a several other differences between Brazilian and Italian politics could also explain potential differences in the results of both papers.

---

[25] Wright and Riker (1989) also find a similar result.

[26] Increases in the bandwidths add precision to the estimates. A linear estimate with a sample that includes all municipalities with more than 50,000 voters finds a TE of 0.843 (se = 0.201), while its quadratic counterpart would be 0.727 (se = 0.254). Larger bandwidths that include even smaller municipalities would lead to unreliable estimates that put excessive weight on the numerous small municipalities.

[27] This could be explained by the payoff of being a mayor in a larger municipality is larger, or that a mayoral campaign larger municipalities generates better opportunities for candidates who wish to increase their visibility for future statewide elections, or that larger cities have a larger pool of potential politicians to run for office.
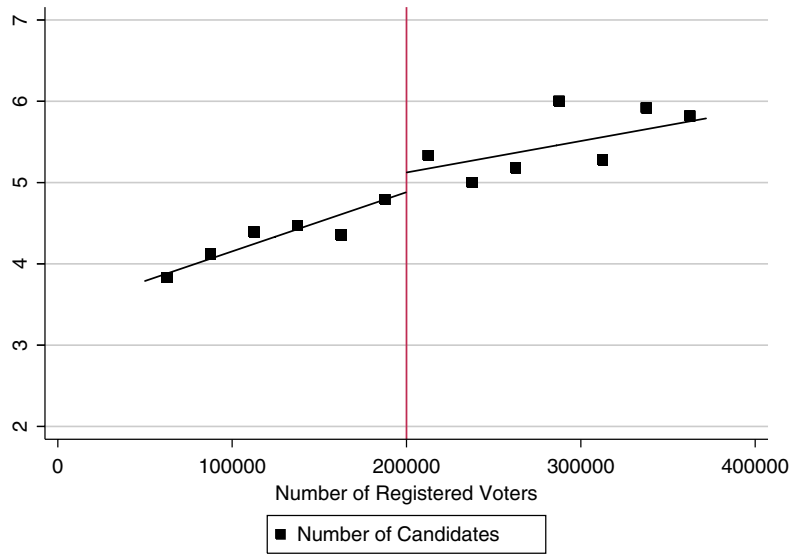
Figure 4. Number of candidates — local averages and parametric fit.

Party Affiliation

This section provides evidence that there is no systematic difference in which parties choose to enter SB and DB mayoral elections close to the threshold. In the period covered in the sample (1996–2008), there were 37 different political parties in activity in Brazil,[28] and 29 of them had at least one top-three candidate in a municipal election in the sample used in the estimations in the previous section.[29]

To check if a particular party is more likely to enter an election under different electoral rules, I estimate the treatment effect of a change from SB to DB on an dummy indicator that takes a value of one if the particular party entered the mayoral race. Owing to space considerations, I focus only on the 15 more relevant parties and present only the effects from a local linear regression with a bandwidth of 50,000 voters.[30]

---

[28] Note there is some amount of party creation and destruction across years. However, any given municipal election year had at least 25 parties in activity.

[29] While different parties are arguably associated with different ideologies in the Federal Congress, party affiliation has little implications to candidate ideology at the municipal level. Ames (2001) discusses the Brazilian party system in detail.

[30] The 15 more relevant parties are defined as those that entered more than 10% of the elections in the sample with a 50,000 voter bandwidth. Reporting TEs with 5 different specifications and

Table 5. Treatment effects on party entry.

Party acronym	Single-ballot mean	Treat. effect	Party acronym	Single-ballot mean	Treat. effect
DEM	0.083	−0.044 (0.067)	PSB	0.250	0.011 (0.122)
PDT	0.417	−0.219 (0.121)	PSDB	0.521	0.094 (0.134)
PFL	0.146	0.078 (0.113)	PSOL	0.208	−0.008 (0.078)
PL	0.146	0.066 (0.106)	PSTU	0.167	0.004 (0.109)
PMDB	0.417	0.177 (0.168)	PT	0.750	0.012 (0.115)
PMN	0.063	−0.007 (0.095)	PTB	0.188	0.009 (0.116)
PP	0.104	0.291 (0.121)	PV	0.146	0.123 (0.126)
PPS	0.208	−0.044 (0.148)	Other Parties	0.042	0.056 (0.066)
Observations	—	175	—	—	175

χ^2-Stat for All Treatment Effects Jointly Significant: 19.40 (p-value = 0.249).

Robust standard errors clustered at the municipality level in parenthesis. The level of observation is a municipal election. The table presents the estimated treatment of a change from SB to DB on a dummy that indicates if the specified party entered the mayoral race. All estimates are based on a local linear regression with a bandwidth of 50,000 voters. All estimates include year effects.

Table 5 presents the results. Parties are referred to by their official acronyms.[31] For example, the *Partido dos Trabalhadores* (Workers' Party)

---

bandwidths (as in Tables 1–3) for all parties would require a table with 185 entries. Moreover, the choice of specification and bandwidth does not affect the qualitative results.

[31] Parties are better known to Brazilian voters by their acronyms than for their name. Ballots, advertisement material, and the media usually refer to parties by their acronym, and not their name. Appendix C lists the parties names and acronyms.
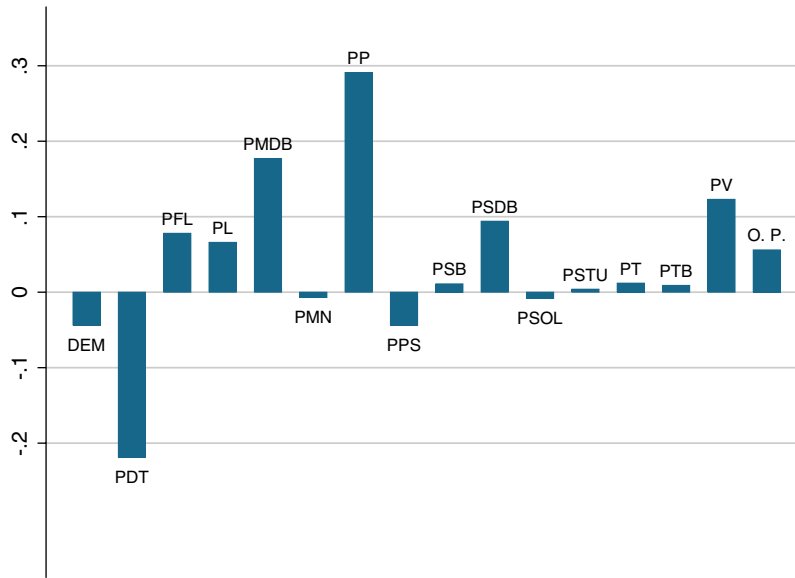
Figure 5. Treatment effects on probability of entry — by party.

is referred to as PT and Table 5 indicates this party entered 75% of the SB elections close to the threshold, and that a change to DB increased the probability that it enters a race by 1.2 p.p. Table 5 also presents the results from an indicator taking value of one if any of the 22 less relevant parties entered the race. The same estimated treatment effects reported in Table 5 are presented graphically in Figures 5 and 6, in which the bars' heights represent the TE sizes. Analogously, Table 5 presents the t-statistics (TE divided by their standard error). Of the 16 TEs presented in Table 5 and Figures 5 and 6, only one is statistically significant at 5% (PP — *Partido Progressista*), an event that can be attributed to random chance. Moreover, the joint test of significance in Table 5 shows that it is not possible to reject the null hypothesis that all the estimated effects are equal to zero at a level of significance below 25%.

   The results in Table 5 and Figures 5 and 6 indicate that there are no systematic differences in which party chooses to have a mayoral candidate under SB and DB elections. Hence, it is possible to rule out the possibility that the effect of electoral rules on vote shares is driven by party entry.
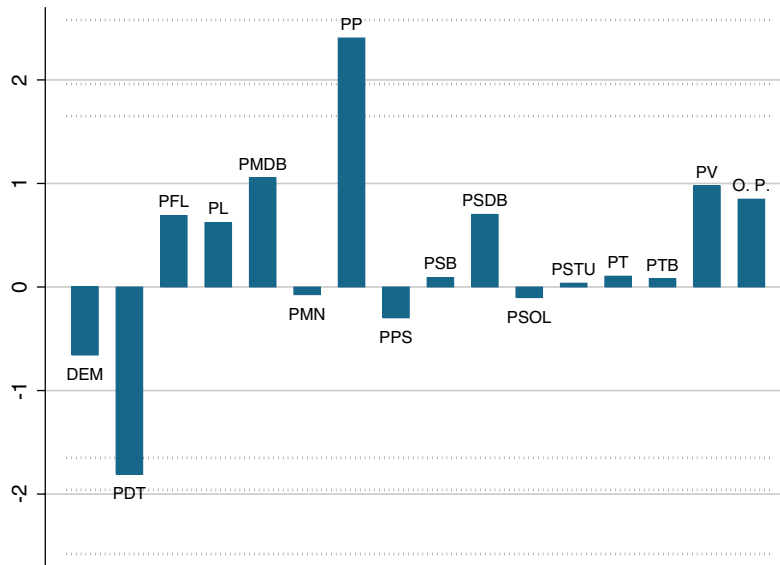
Figure 6. T-statistics of treatment effects on probability of entry — by party.

Dashed lines represent the values for significance at the 10%, 5%, and 1% level.

Candidate Quality

Even if all citizens voted sincerely, it would still be possible to find the treatment effects of the Main Results section if under DB the candidates placed third and lower where of better *quality* than those under SB, where quality is defined as the ability to attract votes. In other words, the results could be explained by systematic differences in candidate characteristics under different electoral rules.

This possibility can be explored by testing if candidates are observably different under SB and DB, which is done by the estimation of treatment effects on observable characteristics of first, second, third, and fourth placed candidates. The available observable characteristics are education and occupation, which are reported by the candidates to the federal elections authority when they register their candidacy.[32]

---

[32] Other information reported by the candidates and made available to the public by the federal elections authority for all election years in the sample is date of birth, marital status, and

This section focuses on three dummy variables measuring candidate characteristics. The first indicates if the candidate has a university degree (*Ensino Superior*). The second one takes a value of one if the candidate has a high school diploma (*Ensino Médio*) or a university degree. The third one indicates if the candidate has a high-skilled occupation, which is defined as medical doctor, dentist, lawyer, manager or entrepreneur.[33]

Note that the relevant issue is the relative quality between third and lower placed candidates and the first and second ones (e.g., the previous results could be potentially caused either by lower quality first and second placed candidates or by higher quality third placed candidates). Hence, it is important to provide the characteristics of different placed candidates under both SB and DB, and Table 6 does so for the first, second, third, and fourth placed ones.

Firstly, Panel A presents the single-ballot means. About 30% of these candidates have a high-skilled occupation, and about 80% finished high school and 65% obtained a university degree. Within variables, there is not substantial variation across candidate position, with the exception that fourth placed candidates seem less likely to attend college.

Panel B then provides the estimated treatment effects on candidate characteristics. Column (1) presents the impact of DB on the probability that the first placed candidate has a high-skilled occupation (first row), a high school degree (second row), and a university degree (last row). Columns (2)–(4) repeat the same exercise for the case of candidates placed in second, third, and fourth position, respectively. Owing to space considerations, all estimates are based on a local linear regression with a bandwidth of 50,000.[34]

Apart from the effect on the probability that the first-placed candidate has a high school degree, none of the estimates are significant at the 10% level. Moreover, the numerical estimates point out that third and fourth

---

gender. However, those variables present a substantial amount of missing information. Moreover, the link between such variables and candidate quality is not as clear as in the case of education and occupation.

[33] Education and occupation information is missing for about 10% of the sample. All missing variables are coded as zero in the construction of the variables. Education and a similarly defined indicator of high-skilled occupation are used as measures of candidate quality for Brazilian mayors by Brollo *et al.* (2010).

[34] Reporting TEs with five different specifications and bandwidths (as in Tables 1–3) for all different placed candidates and characteristics would require a table with 60 entries. Moreover, the choice of specification and bandwidth do not affect the qualitative results.

Table 6. Treatment effects on candidate characteristics.

Candidate characteristic/ dependent variable	Candidate's position in election			
	First (1)	Second (2)	Third (3)	Fourth (4)
Panel A: Single-ballot means				
High-skilled occupation	0.292	0.313	0.311	0.315
High school degree	0.750	0.813	0.822	0.736
University degree	0.646	0.688	0.667	0.552
Panel B: Estimated treatment effects				
High-skilled occupation	−0.054	−0.224	−0.105	−0.117
	(0.125)	(0.151)	(0.123)	(0.128)
High school degree	0.157	−0.046	0.015	−0.004
	(0.078)	(0.087)	(0.079)	(0.106)
University degree	0.191	−0.096	−0.219	−0.074
	(0.117)	(0.144)	(0.167)	(0.178)
Observations	175	175	165	136

Robust standard errors clustered at the municipality level in parenthesis. The level of observation is a municipal election. Panel A presents the single-ballot mean for a dummy indicator of the specified characteristic of the ith most voted candidate on column (i). Panel B presents the estimated treatment of a change from SB to DB on this variable. All estimates are based on a local linear regression with a bandwidth of 50,000 voters. All estimates include year effects.

placed candidates are relatively less likely to have quality indicators under DB. Overall, Table 6 does not support the notion that third and fourth placed candidates are of (relative) higher quality under DB.

To further test if differential candidate quality drives the results, Table 7 presents estimates that replicate those of the first row of Table 1 (the impact of DB on vote share of the third and lower placed candidates) with the addition of controls for candidate characteristics. In order to control for the composition of candidate quality in a flexible manner, I use a set of dummy indicators for all the possible combinations of a quality indicator across the four most voted candidates. Because a given quality variable (e.g., high-skilled occupation) is binary, there are only 16 possible values that its joint

Table 7. Treatment effects with controls for candidate characteristics.

Specification/ bandwidth	Single-ballot mean	Dependent variable: Vote share of third and lower				
		Linear 50,000	Linear 25,000	Linear 75,000	Quad. 50,000	Quad. 75,000
Included controls		(1)	(2)	(3)	(4)	(5)
With university controls	0.155	0.098 (0.043)	0.131 (0.062)	0.072 (0.033)	0.134 (0.056)	0.124 (0.045)
With high school controls	0.155	0.077 (0.037)	0.060 (0.052)	0.049 (0.030)	0.094 (0.051)	0.094 (0.043)
With high-skilled occupation controls	0.155	0.090 (0.040)	0.064 (0.073)	0.062 (0.030)	0.082 (0.067)	0.101 (0.047)
Observations	—	175	81	282	175	282

Robust standard errors clustered at the municipality level in parenthesis. Each figure in the table is from a separate local linear/quadratic regression with the specified bandwidth. The level of observation is a municipal election. The estimated treatment effect is of a change from SB to DB. All estimates include year effects. Details on the dependent variables are presented in the text.

distribution across the top four candidates can take, and hence the controls are a set of 16 dummies that indicate each of these cases.[35]

The first row of Table 7 presents the treatment effects with controls for distribution of university degree status across candidates. The estimates are slightly larger than its counterparts in Table 1, and always significant at the 5% level. In the case of high school status and high-skilled occupation, the estimated effects are of similar magnitude (and statistical significance) to the ones in Table 1. Overall, the results in Table 7 indicate that flexible controls for the observable characteristics of candidates do not affect the result that a change from SB to DB increases the vote share of the third and lower placed candidates.

---

[35] Note that the set of dummies nests the case where an indicator for the quality of each candidate is included, or including average quality (or any other statistic from its distribution).

Candidate Behavior and Unobserved Quality

The final competing mechanism to be addressed is the possibility that changes in candidate behavior are the driving cause for the effect of electoral rules on the vote shares of third placed candidates. Unfortunately, there is scarce data on how intensively a candidate campaigns and which policy positions they adopt, so that directly checking if candidate behavior is similar just above and just below the cutoff is not possible.

This section, however, provides an exercise that indirectly tests if candidate behavior differs under different electoral rules, which exploits that mayoral elections occur simultaneously with elections for municipal legislatures (*Câmara dos Vereadores*). In all municipalities, a voter casts his vote for the municipal legislature at the same time and places that he votes for mayor (in the case of DB municipalities, at the same time of the first round).[36]

Municipal legislature elections are carried under a proportional representation system.[37] As in mayoral elections, the election is at large (a municipality is a single district). Of particular importance is that the electoral rules are exactly the same for cities below and above the 200,000 voter threshold. Hence, systematic differences between the legislative election results in municipalities that have mayoral election under SB and DB should not exist, apart from a possible spill-over effect from the mayoral race to the legislative one.

Given the simultaneous campaigning of mayoral and legislative candidates, it is likely that a coattail effect exists from one to the other: actions that mayoral candidates take to increase their vote share are likely to also have an effect on the vote share of the legislative candidates of the same party. Hence, by checking if the results of legislative elections are affected by the change in mayoral electoral rule from SB to DB, one can test if differences

---

[36] At municipal elections, only mayor and legislature members are elected (i.e., these are the only two votes cast on those election days). Other (state and federal) elected offices in Brazil are chosen in different years from municipal elections, and no plebiscites and referendums were held simultaneously with municipal elections in the period covered in the sample (1996–2008).

[37] Specifically, the system used is open-list proportional representation with seats awarded by the d'Hondt formula. This is the proportional representation system where a voter can cast a vote to individual candidates or party lists. The number of seats awarded to a party is proportional to votes that the party list or party candidates received; however, the votes for which candidate within a party list define which individual gets the seat. Cox (1997) and Ames (2001) provide a more detailed description.

in candidate behavior is the mechanism behind the effect of DB on mayoral electoral outcomes.

A caveat of this exercise is that it requires the assumption that a coattail effect does exist. If the actions of mayoral candidates do not spill-over to legislative candidates, the test described above conveys no information about mayoral candidate behavior (although it does provide a falsification test that adds robustness to the causal interpretation of the RD results).[38] On the other hand, if a spill-over from unobserved mayoral candidate behavior and legislative candidates' performance does exist, no systematic differences in legislature election results in municipalities with SB and DB mayoral races provides evidence that rules out the role of unobserved candidates characteristics in driving the previous results.

Note also that the same argument made above for the case of (unobserved) candidate behavior also applies to their unobserved characteristics. If the mayoral candidate quality of a party affects its performance in the legislature election, the test above also provides evidence on the role of unobserved mayoral candidate characteristics in the previous results.

I estimate the treatment effects on four different municipal legislature electoral outcomes: the share of seats[39] awarded to the party of the elected mayor, the share of seats that are awarded to the most voted (and also the two most voted) parties in the legislature election and the Hirschman–Herfindahl Index (HHI) of party concentration in the elected legislature.[40]

The results are presented graphically in Figure 7 for three of those outcomes. A smooth relationship with no jumps at the 200,000-voter threshold is depicted. The formal counterpart can be seen in Table 8, where the results are mostly close to zero and generally insignificant.[41]

---

[38] Evidence for a coattail effect in Brazilian elections in the sense described above is, to the best of my knowledge, not available. Note that the correlations between vote shares in mayoral and legislative elections do not provide such evidence, as that can be driven by omitted effects and not the actions of mayoral candidates.

[39] Given the proportional representation electoral rule, seat shares and vote shares by party are virtually the same.

[40] The index equals the sum of the squares of the seat shares of each party. Hence it goes from zero (infinite amount of parties, one with each seat) to one (one party has all the seats). The inverse of this measure is commonly referred to as the effective number of parties.

[41] The significant results appear only in the 25,000 bandwidth sample with linear specification and 50,000 bandwidth sample with quadratic specification, which likely implies that an outlier close to the threshold is driving the result.
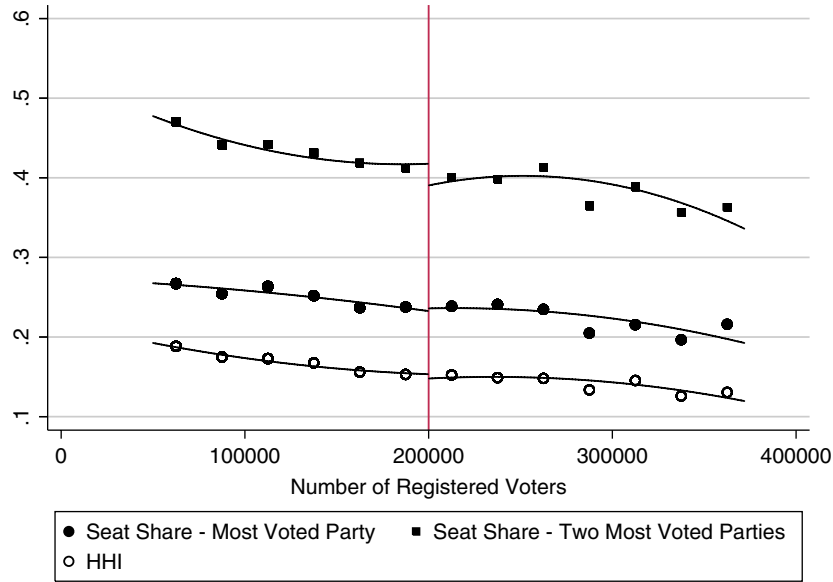
Figure 7. Outcomes of legislature elections — local averages and parametric fit.

Table 8. Treatment effects on municipal legislature election outcomes.

Specification/ bandwidth	Single-ballot mean	Linear 50,000	Linear 25,000	Linear 75,000	Quad. 50,000	Quad. 75,000
		(1)	(2)	(3)	(4)	(5)
Seat share — Mayor's party	0.144	0.017 (0.023)	−0.026 (0.033)	0.018 (0.017)	−0.008 (0.04)	0.015 (0.029)
Seat share — most voted party	0.238	−0.021 (0.02)	−0.063 (0.027)	−0.003 (0.016)	−0.063 (0.031)	−0.033 (0.023)
Seat share — 2 most voted parties	0.412	−0.025 (0.026)	−0.071 (0.034)	−0.022 (0.021)	−0.078 (0.039)	−0.039 (0.03)
HHI	0.153	−0.006 (0.011)	−0.028 (0.015)	0.0003 (0.009)	−0.028 (0.017)	−0.015 (0.013)
Observations	—	175	81	282	175	282

Robust standard errors clustered at the municipality level in parenthesis. Each figure in the table is from a separate local linear/quadratic regression with the specified bandwidth. The level of observation is a municipal election. The estimated treatment effect is of a change from SB to DB. All estimates include year effects. Details on the dependent variables are presented in the text.

The set of results in Table 8 indicates that a change in mayoral electoral rule has no spill-over effect on legislative election outcomes. Non-zero effects would be expected if mayoral candidates changed their behavior in response to the change from SB and DB, with this differential behavior also affecting the performance of legislative candidates.

While these results do not provide direct evidence that rules out the role of mayoral candidate behavior (and unobserved characteristics), it strengthens the case that strategic voting is the driving mechanism behind the effect of DB on third placed candidate vote share. Strategic voting provides an straightforward explanation for the results in mayoral and legislative elections: change in voter behavior is only observed in the outcomes for the election for the office where the electoral rule does change. The alternative mechanism of candidate behavior (and unobserved quality), however, has to deal with the additional complication that no spill-overs to legislative election results occur. In other words, an explanation that relies on mayoral candidate acting differently under DB has also to explain why such differences in behavior have no coattail effects.

Conclusion

The results in this paper can be separated into two components. First, it shows that, in the context of Brazilian mayoral elections, a change from single-ballot plurality rule to dual-ballot lowers the vote share of the top two candidates, to the benefit of the third placed one. The causal validity of this result is likely to hold given the quasi-random assignment of electoral rules generated by their discontinuous assignment across municipalities. The validity of this regression discontinuity design is supported by a number of robustness test.

The above results are consistent with the presence of strategic voter behavior and other alternative explanations. The second component of the paper is a combination of several separate pieces of evidence that suggests that these alternative explanations can be ruled out or require some qualifications. The combination of all the evidence provided can be straightforwardly and parsimoniously explained by strategic voting, while alternative explanations would require more convoluted arguments.

For example, if one would try to explain the effect of electoral rules on vote shares by the mechanism that third-placed parties campaign more intensively under dual-ballots, such explanation would have to address why this more intense campaign is not observed in less contested elections, why the party campaigning more intensively does not put forth a better quality candidate, and how this campaigning does not affect the legislative election outcomes.

In conclusion, although the patterns found in the data make a case for strategic voting, they say very little about the mechanisms that generate the (perhaps self-fulfilling) expectations of which candidates will finish first, second, and third, and how these expectations allow coordination between voters. In the elections used in the estimations, over 150,000 citizens vote. Understanding how such a large number of people coordinate is an useful future direction of research.

Appendix A: Descriptive Statistics

Descriptive statistics is given in Table A1.

Table A1. Descriptive statistics.

Variable	Mean	Std. dev.	Minimum	Maximum
Panel A: Elections with less than 200,000 voters				
(Single-ballot elections — 21,256 observations)				
Vote share — 1st placed candidate	0.555	0.124	0.227	1
Vote share — 2nd placed candidate	0.377	0.106	0	0.500
Vote share — 3rd and lower placed	0.068	0.107	0	0.571
Number of voters	14,076.8	21,243.4	501	199,607
Number of candidates	2.744	1.034	1	10
Panel B: Elections with more than 200,000 voters				
(Dual-ballot elections — 234 observations)				
Vote share — 1st placed candidate	0.505	0.131	0.255	0.942
Vote share — 2nd placed candidate	0.286	0.082	0	0.478
Vote share — 3rd and lower placed	0.208	0.123	0	0.511
Number of voters	605,975.2	1,008,592.0	200,203	8,198,282
Number of candidates	6.218	2.269	2	16

(*Continued*)

Table A1. (*Continued*)

Variable	Mean	Std. dev.	Minimum	Maximum
Panel C: Elections with more than 150,000 but less than 200,000 voters (Single-ballot elections — 113 observations)				
Vote share — 1st placed candidate	0.515	0.122	0.312	0.930
Vote share — 2nd placed candidate	0.327	0.089	0.070	0.498
Vote share — 3rd and lower placed	0.158	0.117	0	0.423
Number of voters	172,468.8	15,461.7	150,206	199,607
Number of candidates	4.540	1.476	2	9
Panel D: Elections with more than 200,000 but less than 250,000 voters (Dual-ballot elections — 62 observations)				
Vote share — 1st placed candidate	0.515	0.149	0.255	0.942
Vote share — 2nd placed candidate	0.290	0.089	0.058	0.477
Vote share — 3rd and lower placed	0.195	0.134	0	0.498
Number of voters	222,690.0	13,677.4	200,203	246,222
Number of candidates	5.177	1.645	2	9

Appendix B: Treatment Effects with Controls

As in a randomized experiment, with a regression discontinuity design consistent estimates of the treatment effects can be obtained without including covariates in the estimations. However, it is common practice to do so for two reasons. First, covariates that are known not to be affected by treatment/control status but are correlated to the outcome variable may increase the precision of the estimates. Second, it provides a robustness check, because the inclusion of the covariates should not affect the size of the estimated treatment effects.

In this section, I repeat the main result of the paper (presented in the first row of Table 1) with different three separate sets of covariates as controls. The first one is the electoral covariates set, which includes the registration rate and the turnout rate. The second set is named economic covariates and includes the *per capita* income, average years of schooling, share of population living in a rural area, and a measure of income inequality (Gini index) in the municipality. Finally, there is a geographical covariates set that includes the municipality's longitude and latitude. All these variables are described in the main results section of the paper.

The results are presented in Table B2. A comparison with the first row of Table 1 shows that the estimates' magnitude and significance are robust to

Table B2. Treatment effects with covariates.

Specification/ bandwidth	Single-ballot mean	Linear 50,000	Linear 25,000	Linear 75,000	Quad. 50,000	Quad. 75,000
		(1)	(2)	(3)	(4)	(5)
Vote share (3rd and lower)	0.155	0.088	0.093	0.069	0.102	0.113
(Electoral covariates)		(0.04)	(0.056)	(0.033)	(0.056)	(0.046)
Vote share (3rd and lower)	0.155	0.085	0.082	0.064	0.098	0.112
(Economic covariates)		(0.038)	(0.053)	(0.031)	(0.055)	(0.044)
Vote Share (3rd and lower)	0.155	0.075	0.066	0.064	0.079	0.101
(Geographic covariates)		(0.038)	(0.058)	(0.031)	(0.056)	(0.044)
Observations	—	175	81	281	175	281

Robust standard errors clustered at the municipality level in parenthesis. Each figure in the table is from a separate local linear/quadratic regression with the specified bandwidth. The level of observation is a municipal election. The estimated treatment effect is of a change from SB to DB. All estimates include year effects. Details on the dependent variables are presented in the text.

a number of different covariates. The size of the standard errors shows, however, that there is not much gain in precision by adding additional controls.

Appendix C: List of Party Acronym and Names

DEM — Democratas
PAN — Partido dos Aposentados da Nação
PC do B — Partido Comunista do Brasil
PCB — Partido Comunista Brasileiro
PCO — Partido da Causa Operária
PDT — Partido Democrático Trabalhista
PFL — Partido da Frente Liberal
PGT — Partido Geral dos Trabalhadores
PHS — Partido Humanista da Solidariedade
PL — Partido Liberal
PMDB — Partido do Movimento Democrático Brasileiro
PMN — Partido da Mobilização Nacional
PP — Partido Progressista
PPB — Partido Progressista Brasileiro
PPS — Partido Popular Socialista

PR — Partido da República

PRB — Partido Republicano Brasileiro

PRN — Partido da Reconstrução Nacional

PRONA — Partido da Reedificação da Ordem Nacional

PRP — Partido Republicano Progressista

PRTB — Partido Renovador Trabalhista Brasileiro

PSB — Partido Socialista Brasileiro

PSC — Partido Social Cristão

PSD — Partido Social Democrtico

PSDB — Partido da Social Democracia Brasileira

PSDC — Partido Social Democrata Cristão

PSL — Partido Social Liberal

PSN — Partido da Solidariedade Nacional

PSOL — Partido Socialismo e Liberdade

PST — Partido Social Trabalhista

PSTU — Partido Socialista dos Trabalhadores Unificado

PT — Partido dos Trabalhadores

PT do B — Partido Trabalhista do Brasil

PTB — Partido Trabalhista Brasileiro

PTC — Partido Trabalhista Cristão

PTN — Partido Trabalhista Nacional

PV — Partido Verde

References

Ames, B. 2001. *The Deadlock of Democracy in Brazil.* Ann Arbor: The University of Michigan Press.

Alvarez, R. M. and J. Nagler. 2000. "A New Approach for Modelling Strategic Voting in Multiparty Elections." *British Journal of Political Science* 30: 57–75.

Bouton, L. 2010. *A Theory of Strategic Voting in Runoff Elections.* Mimeo, Boston University.

Besley, T. and S. Coate. 1997. "An Economic Model of Representative Democracy." *Quarterly Journal of Economics* 112: 85–114.

Bordignon, M., T. Nannicini, and G. Tabellini. 2010. *Moderating Political Extremism: Single vs. Runoff Elections Under Plurality Rule.* Mimeo, Bocconi University.

Brollo, F., T. Nannicini, R. Perotti, and G. Tabellini. 2010. *The Political Resource Curse.* Mimeo, Bocconi University.

Chamon, M., J. M. P. de Mello, and S. Firpo. 2009. "Electoral Rules, Political Competition and Fiscal Spending: Regression Discontinuity Evidence from Brazilian Municipalities." *IZA Discussion Paper n. 4658.*

Cox, G. W. 1994. "Strategic Voting Equilibria under the Single Non-Transferable Vote." *American Journal of Political Science* 88: 608–625.

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

Degan, A. and A. Merlo. 2007. "Do Voters Vote Ideologically?" *Journal of Economic Theory* 144: 1868–1894.

Duverger, M. 1954. *Political Parties.* New York: Wiley.

Engstrom, R. L. and R. N. Engstrom. 2008. "The Majority Vote Rule and Runoff Primaries in the United States." *Electoral Studies* 27: 407–416.

Golder, M. 2006. "Presidential Coattails and Legislative Fragmentation." *American Journal of Political Science* 50: 34–48.

Gonçalves, C. E. S., R. A. Madeira, and M. Rodrigues. 2008. *Two-ballot vs. Plurality Rule: An Empirical Investigation on the Number of Candidates.* Mimeo, University of São Paulo.

Hahn, J., P. E. Todd, and W. Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69: 201–209.

Imbens, G. and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142: 615–635.

Kawai, K. and Y. Watanabe. 2010. *Inferring Strategic Voting.* Mimeo, Northwestern University.

Lee, D. S. 2008. "Randomized Experiments from Non-Random Selection in the U.S. House Elections." *Journal of Econometrics* 142: 675–697.

Martinelli, C. 2002. "Simple Plurality Versus Plurality Runoff with Privately Informed Voters." *Social Choice and Welfare* 19: 901–919.

McCrary, J. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142: 698–714.

Myatt, D. P. 2007. "On the Theory of Strategic Voting." *Review of Economic Studies* 74: 255–281.

Myerson, R. and R. Weber. 1993. "A Theory of Voting Equilibria." *American Political Science Review* 87: 102–114.

Myerson, R. 1999. "Theoretical Comparisons of Electoral Systems." *European Economic Review* 43: 671–697.

Myerson, R. 2002. "Comparison of Scoring Rules in Poisson Voting Games." *Journal of Economic Theory* 103: 219–251.

Mullainathan, S. and E. Washington. 2009. "Sticking with Your Vote: Cognitive Dissonance and Voting." *American Economic Journal: Applied Economics* 1: 86–111.

Osborne, M. J. and A. Slivinski. 1996. "A Model of Political Competition with Citizen Candidates." *Quarterly Journal of Economics* 111: 65–96.

Palfrey, T. 1989. "A Mathematical Proof of Duvergers Law." In *Models of Strategic Choice in Politics*, P. C. Ordershook, ed., Ann Arbor: University of Michigan Press.

Rietz, T. 2008. "Three-way Experimental Election Results: Strategic Voting, Coordinated Outcomes and Duvergers Law." In *Handbook of Experimental Economics Results*, R. Plott and V. Smith, eds., Amsterdam: Elsevier.

Riker, W. H. 1982. "The Two-Party System and Duvergers Law: An Essay on the History of Political Science." *American Political Science Review* 76: 753–766.

Simon, H. A. 1954. "Bandwagon and Underdog Effects and the Possibility of Equilibrium Predictions." *Public Opinion Quarterly* 18: 245–253.

Taagepera, R. 2003. "Arend Lijphart's Dimensions of Democracy: Logical Connections and Institutional Design." *Political Studies* 51: 1–19.

Wright, G. C. 1990. "Misreports of Vote Choice in the 1988 NES Senate Election Study." *Legislative Studies Quarterly* 15: 543-563.

Wright, G. C. 1992. "Reported Versus Actual Vote: There Is a Difference and It Matters." *Legislative Studies Quarterly* 17: 131-142.

Wright, S. G. and W. H. Riker. 1989. "Plurality and Runoff Elections and Number of Candidates." *Public Choice* 60: 155–176.