

NBER WORKING PAPER SERIES

NORMS IN BARGAINING:  
EVIDENCE FROM GOVERNMENT FORMATION IN SPAIN

Thomas Fujiwara  
Carlos Sanz

Working Paper 24137  
<http://www.nber.org/papers/w24137>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
December 2017

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Thomas Fujiwara and Carlos Sanz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Norms in Bargaining: Evidence from Government Formation in Spain  
Thomas Fujiwara and Carlos Sanz  
NBER Working Paper No. 24137  
December 2017  
JEL No. C7,C78,D72

### **ABSTRACT**

Theories of multilateral bargaining and coalition formation applied to legislatures predict that parties' seat shares determine their bargaining power. We present findings that are difficult to reconcile with this prediction. We use data from 2,898 municipal Spanish elections in which two parties tie in the number of seats. The party with slightly more general election votes is substantially more likely to appoint the mayor (form the government). Since tied parties should, on average, have equal bargaining power, this identifies the effect of being the most voted due to a norm prescribing that "the most voted should form government." The effect of being most voted is comparable in size to the effect of obtaining an additional seat. This norm binds behavior even when the second and third most voted parties can form a winning coalition that prefers the most voted not to appoint the mayor. Voters punish, in future elections, second most voted parties that appoint mayors, suggesting that they enforce the norm. We document a similar second-versus-third most voted effect and provide suggestive evidence of similar norms from 28 national European parliaments. A model where elections play a dual role (aggregating information and disciplining incumbents) and different equilibria (norms) can occur is consistent with our results and yields additional predictions.

Thomas Fujiwara  
Department of Economics  
Princeton University  
131 Julis Romo Rabinowitz Building  
Princeton, NJ 08544  
and NBER  
fujiwara@princeton.edu

Carlos Sanz  
Banco de España  
Calle Alcalá 48  
28014 Madrid, Spain  
carlossanz@bde.es

# 1 Introduction

Multilateral bargaining and coalition formation play a role in many economic and political environments, such as mergers, trade negotiations, and conflict. Theories on the topic focus on how coalition payoffs interact with bargaining procedures to determine outcomes. Less attention is devoted to the role of social norms, here understood as informal rules (self-enforcing mutual conventions, beliefs, and expectations) over what is an “appropriate” or “justified” outcome. According to a recent survey, “*the impact of such norms or processes on equilibrium coalition structures is nontrivial, interesting, and largely unexplored*” (Ray and Vohra 2014).

This paper studies bargaining and coalition formation in legislatures, which are key building blocks to understanding political outcomes and public policy. In particular, we study a norm prescribing that the most voted party should hold the executive in a parliamentary system of government (e.g., the prime minister should be a member of the most voted party).

Our evidence comes from Spanish municipalities. Each municipality elects a council by (closed list) proportional representation under the D’Hondt apportionment rule in a single-district election.<sup>1</sup> In its first meeting, the council selects, by *majority* rule, one of its members to be the mayor — a powerful executive position. We focus on 2,898 elections where the *two most voted parties tied in their number of allocated seats*. Moreover, we exploit a regression discontinuity design (RDD) comparing the probability of appointing the mayor between the first and second most voted parties that are few votes apart. Existing theories predict both parties have equal bargaining power. Our results indicate that *the party with slightly more votes is roughly 20 p.p. more likely to appoint the mayor*.

The following example can clarify our argument. Consider a council with 11 seats where three parties ( $A$ ,  $B$ , and  $C$ ) obtained, respectively, 42.0%, 41.9%, and 16.1% of the votes. By D’Hondt rule, parties  $A$  and  $B$  receive five seats each, and party  $C$ , one seat. Theories of legislative bargaining predict that  $A$  and  $B$  should have similar bargaining power and are equally likely to form a government with majority support. Therefore, the probability that party  $A$  appoints the mayor should be equal to party  $B$ ’s. Our findings, however, are that party  $A$  has a 55% probability of forming the government, while party  $B$  has 35%.

Spanish electoral rules do not award an *obvious* advantage to the most voted. The only formal rule treating parties differentially by vote ranks is when defining the “status quo.” If a *majority* of council members cannot agree on a mayor, the leader of the most voted party is appointed.<sup>2</sup> While at first pass it may seem to play an important role, there are multiple reasons this rule is unlikely to explain our results. Perhaps the most compelling is that we find a similar effect of being second (instead of third) most voted (when both parties are tied in seats). There is no rule favoring the second most

---

<sup>1</sup>Proportional representation allocates seats to parties in proportion to their number of votes. In general, exact proportionality is not possible given integer constraints. D’Hondt rule is a commonly used method to allocate seats respecting such constraints.

<sup>2</sup>Only leaders (the higher ranked candidate on the party list) can become mayor, so in our example only three members of the 11-member council can be mayor. The rules are defined in terms of a *majority*: if no party obtains at least  $n/2$  council votes in a council with  $n - 1$  members, the status quo rule assigns the party with most votes in the *general* election becomes mayor — even if such party received less council votes than another party. Section 2 describes the relevant electoral rules in detail.

voted party and hence the status quo rule cannot fully explain our evidence.<sup>3</sup>

This result identifies the existence of a norm—an informal rule—that awards higher ranked parties an advantage in forming government. Intuitively, the RDD isolates a comparison between two groups of parties that should have, on average, equal ex-ante bargaining power. However, one group can be thought of as being randomly assigned the “most voted” label—in a setting where being the most voted should be irrelevant given the (proportional) electoral system. Comparing most voted parties to lower ranked ones in general (i.e., including cases where parties’ vote shares are not almost the same) is potentially confounded by differential number of seats and the other factors that lead one party to outperform the other at the polls. However, these are held constant in our RDDs.

Additionally, by focusing on parties that almost tied in votes, our results indicate that agents respond to variations in *rankings* that provide no additional information conditional on the publicly available continuous variable that fully determines the ranking (vote shares). This adds to evidence of rank-based decision in politics (Anagol and Fujiwara 2016 and Folke, Persson, and Rickne 2015).

The focus on Spanish municipalities is due to the large number of cases where two parties tie in their number of seats. However, we also provide evidence from other contexts suggesting that such norms may not be confined to Spain. Data from 28 national European parliaments indicate that first-placed parties are more likely than second-placed parties to appoint a prime minister. While ties in number of seats rarely occur in this data, the evidence also suggests that the results are driven by an effect of rank itself and not the increased ability to form coalitions. We also briefly discuss anecdotal and survey evidence of related norms in the United States and the United Kingdom.

Our results are difficult to reconcile with existing theories of legislative bargaining and government formation. Their starting point is the number of seats held by voting blocks (e.g., political parties). Given the legislative procedure (majority rule), the number of seats fully determines which coalitions can be formed and thus parties’ bargaining power. In other words, such theories take parties’ number of *seats* as the primitives in a model and thus ignore the role of the votes themselves that lead to these seat allocations. In other words, seat distributions should be sufficient to study bargaining outcomes.<sup>4</sup> Seat allocations are also the focus of empirical work. For example, Gamson’s Law is the empirical regularity that coalition governments distribute cabinet positions in proportion to member parties’ contribution of *seats*. Structural models of government formation also take seat allocations as their starting point. Indeed, datasets used in this literature, such as the *European Representative Democracy Data Archive*, usually contain information only on parties’ seat allocations (and not their general election votes).<sup>5</sup>

---

<sup>3</sup>Sections 2 and 6 discuss additional reasons why the status quo rule cannot explain our results. One relevant issue is that mayors require continuous support from a majority during the term. Hence, in our example, parties *B* and *C* can appoint, say, a *B* mayor anytime they agreed to do so—making it unlikely the status quo binds. We also later describe evidence that voters punish parties that deviate from the norm—it is unclear why a status quo rule would generate this behavior.

<sup>4</sup>This applies to cooperative theories of coalition formation (e.g., the core, stable set, or bargaining set) and to non-cooperative theories. Ray and Vohra (2014) survey coalition formation and Laver (1998) surveys government formation in particular. Examples of noncooperative theories of legislative bargaining and government formation are Baron and Ferejohn (1989), Austen-Smith and Banks (1990), Laver and Shepsle (1990), Baron (1991, 1993), Merlo and Wilson (1995), Morelli (1999), and Snyder Jr, Ting, and Ansolabehere (2005).

<sup>5</sup>Gamson’s Law is discussed further in Laver (1998) and Carroll and Cox (2007). Examples of structural estimation of government formation are Merlo (1997) and Diermeier, Eraslan, and Merlo (2003).

While our results can be understood as a “tie-breaking rule,” it is perhaps surprising that such a rule binds behavior given the high stakes involved. The mayor is a key powerful position. Furthermore, we provide evidence suggesting that mayors from the most voted parties do not “share power” with the second most voted party through deputy mayor allocations.

Moreover, in a substantial number of cases, the norm’s prescription binds behavior even when it goes against the perhaps most natural factor determining legislative bargaining outcomes: programmatic affinity between parties. In particular, the effect of being the most voted party is of similar magnitude when we restrict attention to cases where the most voted party is the main right-wing party (*Partido Popular*—PP), while the second and third most voted parties are, respectively, the main left-wing party (*Partido Socialista Obrero Español*—PSOE) and its common leftist ally (*Izquierda Unida*—IU). This implies that, even though the two left-wing parties have a combined majority that could appoint the mayor, the norm binds frequently and the right-wing party appoints the mayor instead.<sup>6</sup>

To gauge the magnitude of the effects of the norm, we compare it to the effect of obtaining one additional seat, which can also be identified using a similar RDD. The effect of having a plurality (but not a majority) of seats is only slightly larger than the effect of being the most voted party but tied in number of seats. This suggests that the importance of the norm we study is comparable to that of previously studied determinants of bargaining outcomes.

We then discuss the possible mechanisms that can drive our results. We start by presenting a simple model of political accountability,<sup>7</sup> in which elections have two roles: information aggregation and incumbent disciplining. Elections aggregate dispersed information about an uncertain state of the world. Thus, *after* an election, voters update beliefs about which party they prefer would appoint the mayor. However, parties’ representation in the council is already set at this point, and bargaining over mayoral appointments can be based on rent allocations that ignore voters’ interests. This creates a conflict between voters and parties and the need for the former to discipline the latter.

The model has multiple equilibria, which can be interpreted as different norms (self-enforcing rules of behavior) that voters can adopt.<sup>8</sup> A norm that matches our results, where voters punish second most voted parties that appoint the mayor, constitutes an equilibrium. In this case, the most voted party becomes more likely to appoint the mayor—even though the model assumes a bargaining procedure that treats the first and the second most voted parties interchangeably. The norm also maximizes voters’ expected welfare and thus has instrumental value.

We provide two pieces of evidence consistent with the model. First, voters appear to enforce the norm by punishing parties that deviate from it. We document this using a triple-differences strategy that leverages variation across time, whether a party was barely the first or second most voted (but tied in number of seats), and whether it appointed the mayor. Second most voted parties that appoint the mayor lose votes in the next election, compared to most voted parties that appoint the mayor.

---

<sup>6</sup>We also document that programmatic affinities are, on their own, predictive of outcomes. In the overall sample, when the PSOE and IU hold a combined majority of seats, it is likely that one of the parties appoints the mayor.

<sup>7</sup>The model is similar in spirit to Barro (1973), Ferejohn (1986), and Persson, Roland, and Tabellini (1997).

<sup>8</sup>Persson, Roland, and Tabellini (1997) also discuss how different equilibria in retrospective voting models can be interpreted as different norms.

Additionally, in a survey, 55% of Spaniards agree that “*it is more democratic that the most voted party forms the government, even if that party does not have an absolute majority of the votes*” (El Pais 2015). Second, the model predicts a specific pattern of heterogeneous effects: the effect of being most voted is stronger when the vote share of the third-placed party is larger. Alternative explanations for our results—for example, the status quo rule or the norm arising as an “agreement” among parties in repeated bargaining—do not naturally lead to these.

This paper is related to five broad strands of the literature. First, as previously discussed, our result is difficult to reconcile with theories of bargaining and coalition formation. It thus suggests the importance of a relatively unexplored determinant of bargaining outcomes. Similarly and also as previously discussed, empirical analyses of legislative bargaining and government formation also take seat allocations as their starting point and thus cannot speak directly to our results.

Second, the results are relevant for comparative politics and the design of electoral systems. In particular, the norm we study can add first-past-the-post considerations to proportional systems. For example, Lizzeri and Persico (2001) associate proportional representation with parties maximizing vote shares (which translates proportionally to power) and plurality rule with parties maximizing the probability of being most voted (a winner-takes-all contest). Our results suggest that incentives under proportional representation may be more similar to those under plurality rule than previously acknowledged, affecting the policy and welfare consequences of electoral rule design. Relatedly, some countries (e.g., France, Greece, Italy, and Portugal) award a seats’ premium to the most voted party in their proportional representation systems. The norm we study can generate equivalent *de facto* premia even without such explicit rules.<sup>9</sup>

Third, we contribute to a growing literature that brings “behavioral” insights and non-standard decision making to political economy contexts.<sup>10</sup> It is particularly related to models where voters see being a “winner” as having value in itself (Callander 2007, Callander and Wilson 2008, and Agranov, Goeree, Romero, and Yariv 2017) and evidence that rank-based decision making affects political outcomes (Anagol and Fujiwara 2016 and Folke, Persson, and Rickne 2015). As previously discussed, by focusing on parties that almost tied in votes, our results indicate that agents respond to variations in rankings that provide no additional information conditional on the publicly available continuous variable that fully determines the ranking (vote shares).<sup>11</sup>

Fourth, our results bring nuance to the “puzzle of nominal versus real bargaining weights” since variations in seat shares that do not generate “real” changes in bargaining power affect allocations. One manifestation of this puzzle relates to the previously discussed Gamson’s Law, which can also be interpreted as a norm of how parties share power, related but separate from the one we study in this

---

<sup>9</sup>The consequences of proportional versus plurality rules are also studied by Persson and Tabellini (2005).

<sup>10</sup>For example, Bendor, Diermeier, and Ting (2003), Bendor et al (2011), Bisin, Lizzeri, and Yariv (2015), Degan and Merlo (2011), Passarelli and Tabellini (2017), and Ortoleva and Snowberg (2015).

<sup>11</sup>Pope (2009), Kuziemko et al (2014), and Hartzmark (2015) study rank-based decision-making in hospital choice, preferences for redistribution, and investment decisions, respectively. The salience of ranks can also be understood as a case of “limited attention” in which decisions are simplified by processing only a subset of available information. Bordalo, Gennaioli, and Shleifer (2013) provide a theory of salient decision-makers. Chetty, Looney, and Kroft (2009) and Finkelstein (2009) study salience issues in the case of taxation, Brown, Hossain, and Morgan (2010) in shipping costs, Anagol and Kim (2012) in mutual funds’ fees, and Lacetera, Pope, and Sydnor (2012) in car purchases.

paper.<sup>12</sup>

Fifth, it also speaks to the literature on why legislatures operating under majoritarian rule support procedures that restrict the majority’s power.<sup>13</sup>

The next section briefly describes the context of Spanish municipal elections and our data. Section 3 presents our main results, which document the existence of the norm. Section 4 describes the theoretical framework and Section 5 the additional evidence it suggests. Section 6 discusses alternative explanations. Section 7 presents evidence from national European parliaments and Section 8 concludes.

## 2 Context and Data—Spanish Municipal Governments

**Electoral Rule and Government Formation Procedure.** Spanish national law regulates how municipal governments are elected and formed. As in a parliamentary system, there are two steps in appointing the executive. First, voters elect a municipal council in a general election. Second, the members of the council elect one of its own to be the mayor.

General elections occur simultaneously in all municipalities every four years. Councils (*concejos*) are elected by proportional representation in single-district (at large) elections. The number of seats in the council is always odd and determined as a function of the municipality’s population one year before the election, as shown in Table A1. Each party presents a ranked list of candidates, determined ahead of the election. On election day, each voter picks one of the party-lists. The conversion from the votes to the seats obtained by each party follows the D’Hondt rule. Political parties must also obtain at least 5% of the votes to receive seats.<sup>14</sup>

In the first council meeting after the election, councilors elect a mayor (*alcalde*). The leaders of each party (the candidates that were ranked first in the pre-determined party-lists) are all eligible to become mayor. Each councilor can vote for one of the councilors or abstain. If one of the candidates obtains a majority (more than 50%) of the votes, then she is appointed mayor. If no candidate obtains a majority, a status-quo rule dictates that the leader of the most voted party in the general election is appointed mayor. Section 6 discusses why this status-quo rule is unlikely to drive our results.<sup>15</sup>

The mayor can be replaced at any moment throughout the term, by two different mechanisms. One is a censure motion (*moción de censura*): a proposal to both remove the current mayor and appoint another councilor as mayor. This requires approval by a majority of the council (and the acceptance

---

<sup>12</sup>See Warwick and Druckman (2001) and Frechette, Kagel, and Morelli (2005). The puzzle is best explained by an example. In a 9-seat legislature where a majority (5 votes) is needed to form a government, a 4-4-1 seat distribution does not differ from a 3-3-3 in terms of “real” bargaining power (all parties have equal “real” power in both cases, since two out of three parties are needed for a coalition). However, if one finds that the smallest party receives less surplus in the 4-4-1 case, that implies that “nominal” considerations matter.

<sup>13</sup>Krehbiel (2004) and Diermeier, Prato, and Vlaicu (2015).

<sup>14</sup>The system described here refers to municipalities with more than 250 inhabitants. Municipalities with 250 or fewer inhabitants use a different, open-list electoral system—Sanz (2017a) and Sanz (2017b) estimate the effects of the change in electoral systems at the 250-inhabitant threshold on voter turnout and women’s representation, respectively. In this paper, we exclude municipalities that do not follow the proportional representation system.

<sup>15</sup>Note, in particular, that we find effects of being second most voted versus third most voted, and there is no rule that awards an advantage to the second most voted party. If no candidate receives a majority and two or more parties obtained the exact same number of votes in the *general election*, then a lottery is run among the tied parties (in 1979 and 1983, such ties in *general election* cases were decided by appointing the oldest party leader). Exact ties in votes in the *general election* are very uncommon, and its few occurrences are deleted from our sample.



by the proposed new mayor). Only one censure motion can occur per term. The other is a motion of no confidence (*cuestión de confianza*), which is proposed by the mayor in certain cases requiring approval of the council (e.g., approving a budget). The number of votes required for the mayor to lose the motion depends on the context in which it is proposed (e.g., in the context of a budget vote, the mayor loses if there are more nays than yeas). If the mayor loses the motion, then the city council elects a new mayor according to the same rules that are used for electing the mayor for the first time (with the exception that the candidate from the party of the removed mayor is now the next person in that party-list). Note that there is no circumstance that leads to an “off-schedule” or “snap” general elections: citizens only vote in municipal elections every four years.<sup>16</sup>

This combination of rules implies mayors not only need to obtain the support a majority to first get elected, but also must keep the support of that majority throughout the term, as it is straightforward for a different majority to appoint a new mayor. In practice, in 97% of cases a mayor from the same party stays in office for the entire term. This number is 89% in the cases where the top two parties tie in number of seats, and 88% when they do so under a small (below 1 p.p.) vote share difference.

**Municipal Governments and Mayors in Spain.** Municipal governments manage approximately 15% of the Spanish public expenditure (6% of GDP). Spanish law dictates which services must be provided by municipal governments (with more populous municipalities having more responsibilities). Municipal governments also collect their own taxes on residential properties, businesses, vehicles, and collect fees and user charges. The total tax and fee collection by municipal governments is approximately 4% of the Spanish GDP. Hence, municipal-level policymaking has substantial consequences to voters.<sup>17</sup>

Mayors are the “*the center of gravity of political life in the municipality*” who “*by law holds the most important executive functions and exercises leadership in municipal politics*” (Vallés and Brugué 2001). They have a central role in running the government by chairing council meetings and appointing and dismissing cabinet members and staff. They have substantial control over determination and allocation of expenditures, since they prepare municipal budgets and approve construction processes. Indeed, Spanish municipal governments exemplify a cases of strong executive power (Sweeting 2009) and have been described as “*municipal presidentialism*” (Magre-Ferran and Bertrana-Horta 2005).

**Data.** Our sample comprises all municipal elections in Spain since the restoration of democracy in 1975. Elections have occurred in exact four-year intervals since 1979. The source is the *Instituto Nacional de Estadística* (INE). We exclude from the sample municipalities that do not use the proportional representation system (i.e., those with less than 250 inhabitants). Our sample is based on the councils elected in the 1983-2011 elections. The sample covers 37,122 elections from 5,993 different

---

<sup>16</sup>If a mayor resigns, is found guilty of a crime, or dies, a new mayor is elected according to the procedure used in the beginning of a new term.

<sup>17</sup>Bagues and Campa (2017) describe the role of municipal governments in further detail. All municipal governments must provide lighting, graveyards, refuse collection, street cleaning, and water and sewer. Larger municipalities must provide social services and education. Some small municipalities may choose to provide services (e.g., childcare) that are not mandated by law.



municipalities. 2,898 elections have the first and second most voted parties tied in number of seats.<sup>18</sup>

We observe the party affiliation of mayors. Unfortunately, we do not observe her supporting coalition. Neither the identity of the members or parties that voted for a given mayor, or the vote count of the election for mayor within the council, are recorded by the INE. Information about the allocation of cabinet positions within municipalities is also unavailable.<sup>19</sup>

**Characteristics of Municipalities Identifying the Results.** There are 438 elections in which the two most voted parties tie in seats and the vote share difference between them is below 1% of the total. In these cases, 90% have councils such that a majority requires support from two of the three most voted parties. This includes both cases where no more than three parties received representation, or cases where the fourth placed party cannot be pivotal in creating a majority (e.g., a 11-seat legislature with a 4-4-2-1 vote division). Therefore, the vast majority of the councils in our sample can be thought of as essentially three-party councils. The coalition formation game that approximates this context is thus one where any two out of three players can form a coalition that allocates payoffs: a “three-player majority game.”<sup>20</sup>

These 438 councils where two parties tie in number of seats and their vote share difference is below 1 p.p. are also relatively small (79% have 13 or fewer legislators) and have the first and second most voted parties “almost tying,” on average, with 36.5% of votes and 39% of seats and the third most voted obtaining vote (seat) share of 17.7% (16.7%).

Given that mayors may not necessarily serve the entire four-year term, we define our main outcome—whether a party “appointed the mayor” or not—as a dummy taking value equal to one if the mayor that spent three quarters of the term in power during the term belongs to that party.<sup>21</sup>

### 3 Main Results - Spanish Municipal Governments

#### 3.1 Empirical Strategy

Our main results are implemented using a regression discontinuity design (RDD). A key aspect is that the sample is restricted to i) only cases where the first and second most voted parties have the exact same number of seats, and ii) only include the first and second most voted party. This sample has  $2,898 \times 2 = 5,796$  observations.

Define  $v_{mt}$  as the vote share of the most voted less the vote share of the second-placed in municipality  $m$  at election year  $t$ . We define the running (forcing) variable, which varies at the party-

---

<sup>18</sup>The 2015 election is not included since the term is still in progress (and mayoral appointments may change). The 1979 election is not included since we use lagged values in placebo tests. We exclude from the sample a small number of cases with missing election results or mayor’s party identity. We also exclude the few elections where there is an exact tie in general election votes between the first and second most voted (or second and third most voted) parties.

<sup>19</sup>We do observe the party affiliation of deputy mayors. This information is used when addressing how power is shared across parties, and described further on Section 3.2.

<sup>20</sup>See Osborne and Rubinstein (1994) (in particular, chapters 13 and 14) for a discussion of majority games.

<sup>21</sup>We condition on spending at least three quarters of the term to exclude cases in which two parties share the term in two halves (which occur rarely) and so that our definition is not sensible to cases where mayors spend a very short (weeks) in office. However, the effects are virtually the same if we define the outcome as being mayor for more than any other party, being the first mayor to be appointed, or serving the full term (89% of cases).

municipality-year level, as follows:

$$x_{pmt} = \begin{cases} v_{mt} & \text{if } p \text{ is the most voted} \\ -v_{mt} & \text{if } p \text{ is second most voted} \end{cases}$$

where  $p$  refers to a political party.

Therefore, if  $x_{pmt} > 0$ , then party  $p$  has the most votes (“first-place”), and it has the second most votes otherwise. Let  $y_{pmt}$  be an outcome (e.g. appointing the mayor). The effect of having most votes is given by  $\lim_{x_{pmt} \downarrow 0} E[y_{pmt}|x_{pmt}] - \lim_{x_{pmt} \uparrow 0} E[y_{pmt}|x_{pmt}]$ , which can be estimated by a local polynomial regression:

$$y_{pmt} = \theta_0 + \theta_1 \cdot 1\{x_{pmt} > 0\} + g_0(x_{pmt}) + g_1(x_{pmt}) \cdot 1\{x_{pmt} > 0\} + \epsilon_{pmt} \quad (1)$$

using observations within a given bandwidth around the threshold.  $g_0$  and  $g_1$  are polynomials estimated separately on each side of the cutoff. Thus,  $\theta_1$  captures the effect of being the most voted (instead of second most voted) in an “almost tied” election in which the two most voted parties obtained a similar share of the votes (and the exact same number of seats). If the most voted and the second most voted are equally likely to appoint the mayor, then  $\theta_1 = 0$ .

Our baseline estimates use a linear specification ( $g_0 = g_1 = x_{pmt}$ ), as suggested by [Lee and Lemieux \(2010\)](#), and use the [Imbens and Kalyanaraman \(2012\)](#) procedure to calculate the optimal bandwidth. We also provide results regarding robustness to different bandwidth choices and polynomial orders. Standard errors are clustered at the municipality level.

There are three noteworthy aspects of this setup. First, variables that do not vary across parties within an municipality-year (e.g., the vote share of the third-placed party or election day weather) will, mechanically, be distributed symmetrically around (and without a possible “jump” at) the RDD threshold. For each election, both a first and second placed party enter the sample in symmetric fashion (one has  $x_{pmt} = a$  and the other  $x_{pmt} = -a$ ). Intuitively, the variation that identifies the results comes from comparing parties within an election and the small vote difference that reshuffles which one is labeled the most voted.<sup>22</sup>

Second, conditioning which observations enter the sample by a variable that varies only at the election or municipality level will not affect the “internal validity” of estimates (in the sense of creating an RDD sample is that is balanced in predetermined covariates around the cutoff). For example, restricting the sample on only the cases with ties in seats should not create imbalances in covariates around the cutoff since the “tie in seats indicator” varies at the election level. A similar logic applies to conditioning the sample on, say, the vote share of the third placed. Third, while each election enters the sample twice, this “double-counting” of elections does not artificially affect our standard errors, since they are clustered at the municipality level.

---

<sup>22</sup>Figure A1 presents the distribution of observations, demonstrating this symmetry.

## 3.2 Identifying the Norm: Effects of Being Most Voted

**Main Result.** The graphical representation of our main result is Figure 1a, which plots the probability that a party appoints the mayor against the vote share difference between the first and second most voted parties (recall that parties that placed third and lower are excluded from the sample). The 5,796 observations are aggregated into bins of one p.p. width of the running variable ( $x_{pmt}$ ), and the local averages for each bin are plotted. The solid lines are from a quadratic polynomial based on the original (unbinned) data and fitted separately on each side of the cutoff. As mentioned above, the graph uses data only from elections where the first and second most voted parties tied in seats.

A clear jump at the cutoff is visible in the Figure 1a (the graphical counterpart to  $\theta_1$ ). It indicates that the second most voted party appoints the mayor 33.6% of the time, while the first most voted party does so 53.9% of the time. This is obtained when focusing in cases where both parties have the exact same number of seats and first and second placed parties are, on average, comparable in every characteristic, given the small vote margin and the RDD rationale.<sup>23</sup>

A perhaps surprising pattern in Figure 1a is that an upward slope is not observed. This indicates that, conditional on the rank of vote shares, higher vote margins for the most voted parties are not associated with increased probability of appointing the mayor. Such slopes must be interpreted with caution, as there are both compositional effects and “omitted variables” that can drive the relationship between the running variable and mayoral appointments.<sup>24</sup>

Panel A of Table 1 shows the equivalent regression results. A second most voted party that almost ties in votes with the most voted (but has the same number of seats) has a 35% chance of appointing the mayor (the “2nd-Place Mean”, which is the estimated  $\theta_0$  from equation 1). Column (1) indicates that the most voted that almost tied in seats is 19 p.p. more likely to appoint the mayor (the estimated  $\theta_1$ ). Both figures are based on using only the 2028 observations from elections where the top-two parties are only 2.32% of the total votes apart (the optimal bandwidth). Column (2) compares the average outcome for the first and second most voted parties that are only 1% of total votes apart and finds a similar effect. Even when focusing on the 46 observations from even closer cases (bandwidth below 0.1%), the estimated effect is 0.522 (s.e.=0.167). Both the linear specifications are robust to the choice of bandwidth (Figure A4). Using the full sample and a quadratic or cubic polynomial yields similar results, shown in columns (3) and (4). These results are statistically distinct from zero at levels well below 1%.

The results are also similar if different definitions of the outcome are used. Table A2 replicates Panel A of Table 1 when defining “appointing the mayor” as either serving the entire term, for a length longer than the other parties, or appointing the initial mayor. Panel E of Table A2 restricts the

<sup>23</sup>In such elections (where the vote share difference is below 1 p.p.) the third-placed party appoints the mayor 3.7% of the time, and the fourth and fifth placed do so 0.5% of the time each. In the remaining 8% of cases, no party appointed a mayor that served for at least three quarters of the four-year term.

<sup>24</sup>Such compositional effects and “omitted variables” are constant when comparing cases around the cutoff. A possible compositional effect occurs since municipalities with smaller councils are less likely to be further away from the cutoff (e.g., it is impossible for two parties that have a 8 p.p. vote share difference to have the same number of seats in a 21-member council, but it is possible in a 7-member council). However, Figure A2 replicates Figure 1a for specific council sizes and indicates the lack of slope is not driven by such compositional effect. Similarly, Figure A3 replicates Figure 1a for specific council seat compositions.

sample to the cases where two of the top three most voted parties are needed to form a majority. In all cases, the estimated effects are similar in magnitude to those in Panel A of Table 1. Lastly, Figures A2 and A3 replicate Figure 1a for subsamples restricted to specific council sizes and seat configurations, indicating that no specific type of council uniquely drives the result.

**Covariate Balance.** To assess the validity of the RDD, Figure 1b repeats the exact same exercise from Figure 1a, but with the lagged outcome on the y-axis. It thus plots whether the party appointed the mayor in the *previous* term against their current vote share difference. The absence of a jump at the cutoff indicates that close first and second most voted parties are equally likely to be the incumbent mayor. Figure A5 repeats this exercise for party identity, showing that neither of the two main national parties, the PSOE and the PP, are more likely to finish in first place in a close election. Panels B and C in Table 1 present the analogous regression results. The point estimates are close to zero and statistically insignificant. These results indicate that incumbent mayors or the major parties are not able to manipulate election results and become the most voted in close elections. Recall that, as discussed above, for any variable that does not vary across parties within an election (e.g., the number of parties running, average municipal education, or the vote share of the third placed party), there is perfect balance by construction. Similarly, there cannot be “bunching” of municipalities around the cutoff.<sup>25</sup>

**Does the norm bind when the second and third most voted are aligned?** To study how the norm interacts with considerations based on programmatic lines, we focus on the cases in which the right-wing PP are the left-wing PSOE are the most voted parties tying in seats, and the left-wing IU is the third most voted. In such cases, the combination of left-wing parties (PSOE and IU) has a majority of seats and is able to appoint one of their leaders as mayor, regardless of whether the PP is the most voted or not. However, our results suggest this is not the case given the norm we study.<sup>26</sup>

The red triangles on Figure 3 replicate the exercise of Figure 1a, but restrict the sample to only cases where the observation regards the PSOE in an election where the PP is the other top two most voted party and the IU is the third most voted party. Hence, the jump at the cutoff indicates that, when the PSOE is barely the second most voted, it appoints the mayor 55% of the time. If the PSOE is the most voted, it appoints the mayor 80% of the time. Similarly, the blue circles indicate that when the PP is the second most voted by a close margin, it appoints the mayor only approximately 10% of the time, however, when it the most voted, it appoints the mayor almost 35% of the time. Table A3 provides the corresponding table.<sup>27</sup>

This result is likely surprising since it makes clear that the PSOE is, overall, much more likely to

<sup>25</sup>For every party with forcing variable  $x_{pmt} = a$ , there is also a party with  $x_{pmt} = -a$  from the same election. Hence there cannot be bunching in the distribution of municipalities by the forcing variable, which must also be symmetrical. This can be seen in Figure A1 which shows how many observations are in each bin of Figure 1a.

<sup>26</sup>97% of the councils in our sample are such that two out of the top three parties are needed for a majority. The antagonism between PP and PSOE is a clear feature of Spanish politics (e.g., both parties never formed a coalition at the federal level and, to our knowledge, only did so once at the regional level, under unusual circumstances in the Basque Country).

<sup>27</sup>The estimates on Table A3 are obtained by estimating equation (1) restricting the sample to cases there the party  $p$  is the PSOE (Panel A) or PP (Panel B) and, in both cases, the IU is the third place and the PP and PSOE are the two most voted parties tying in seats. Note the subsample is not defined by which party is the most voted, and hence it retains a balanced RDD.

appoint the mayor than the PP when the IU is the third place (red triangles well above blue circles in Figure 3). The cases on each side of the RDD cutoff are both in which the PSOE and IU have a combined majority. However, it is still the case that a small difference in vote shares that awards the PP the “most voted label” is enough to substantially increase the chance it appoints the mayor. In other words, the norm “bites” even in the cases with strong alignment between the second and third most voted parties can form a coalition that would appoint a mayor of their own.

A particular case in our sample can illustrate the argument further. The results of Olivenza’s 2011 election was the PP obtaining 2912 votes and 7 seats; the PSOE, 2886 votes and 7 seats; and the IU 1376 votes and 3 seats. Given the 17-member council, the PSOE and IU could appoint one of their leaders as mayors. However, the PP appointed the mayor. The IU leader justified their decision of not supporting the PSOE to the media by stating it needed to accept “*the decision of the people*” and “*what democracy has said*,” even though “*it hurts me*” that we will have a government “*from the right*” (Europa Press 2011). The surprising aspect is that the “*decision of the people*” is based on only 26 votes out of more than 7,000. Moreover, in a proportional representation system that does not formally reward the most voted, a majority of voters preferred the left-wing parties, but the “most voted label” seemed to matter beyond that.

**Are mayoral appointments just symbolic?** One possible interpretation of the results is that mayoral appointments are mostly symbolic. For example, parties could bargain over rent allocations and policy decisions based on their seat distributions and ability to form majority coalitions, and once those are decided, simply appoint the most voted party the mayor symbolically.

There are five reasons this interpretation is unlikely. First, as discussed in Section 2, mayors are dominant figures that exercise substantial personal discretion over policy once appointed. Second, it is not clear why the norm would arise if it was just symbolic (e.g., why parties in a coalition would not appoint mayors for half a term each). Third, if the appointment is symbolic, it is unclear why voters would punish parties that deviate from it, as discussed in Section 5. Fourth, the stakes of appointing a mayor presumably become larger as the size of the municipality grows, both because it controls a larger budget and because larger municipalities must, by law, provide additional public services. Figure A2, however, indicates that RDD “jumps” are larger in municipalities with larger councils, which are also those with more inhabitants (Table A1).<sup>28</sup>

Fifth, it does not appear that mayors share power with other parties. While we do not observe cabinet allocations across municipalities, we do observe the appointment of deputy mayors (*tenientes de alcalde*), which are the second most visible position in municipal government. There may be more than one deputy mayor in a municipality, in which case they are ranked. In cases of absence or illness of the mayor, the (first) deputy mayor takes over her duties.<sup>29</sup>

Formally, the mayor has the discretion to choose which of the elected councilors are appointed deputy mayors. However, it is possible that deputy mayor positions and the mayoral appointment

<sup>28</sup>There are, of course, other characteristics that are correlated with municipality size that could affect effect sizes. For example, larger municipalities are more likely to have a local media outlet that makes council decisions salient to voters.

<sup>29</sup>If the first mayor is also not available, the second deputy mayor takes charge, and so forth. The mayor has discretion on the number of appointed deputy mayors. In municipalities with more than 5,000 inhabitants, the number of deputy mayors must be between one and one third of the council size. In those with less than 5,000, appointing a deputy mayor is optional. The average municipality in our sample of ties in seats has 2.6 deputy mayors (s.d.=1.6).

are bargained over by parties. Hence, one could expect that the “effect of most voted” on mayoral appointments is offset by a *negative* effect on appointment of deputy mayors (i.e., most voted are more likely to appoint mayors and the second most voted then appoint most deputy mayors).

Figure A6 repeats the exercise of Figure 1a, using instead the share of deputy mayors appointed by the party and a dummy for the party appointing *all* deputy mayors. It also provides placebo tests based on lagged outcomes. Table A4 provides the corresponding estimates (as in Table 1).<sup>30</sup> The estimated effects are all *positive* and sizable. Hence, being the most voted party increases the chance of appointing the mayor and the share (or having all) of the deputy mayors. This is the opposite of what the hypothesis that the two most voted parties obtain equal amount of power would predict.

**Which types of coalitions drive the effect?** While we cannot directly observe which parties are supporting the mayor, we can indirectly study whether coalitions by the most voted and second most voted, or between the most voted and third-placed party, drives the effects. Although we cannot observe the supporting coalitions, we leverage previously discussed evidence that left-wing PSOE and the right-wing PP are unlikely to support each other’s governments. Figure A7 replicates Figure 1a, but separately plotting the cases where i) both the PP and PSOE are the two most voted parties, ii) only one of them was the amongst the two most voted parties, and iii) neither are amongst the two most voted parties. Table A5 provides the corresponding estimates. The effects are similar (and not statistically distinct) in all three cases. This suggests it is not the case that the norm is driven systematically by coalitions where the second most voted (or third placed) party supports the first most voted.

### 3.3 Effect of Being Second (Instead of Third) Most Voted

The exercise of the previous subsection can also be applied to estimate if being labeled the second most voted, instead of third most voted, also has an effect on the probability of appointing a mayor. To do so, we redefine the sample such that i) it only includes elections where the *second* and *third* most voted party obtained the *same number of seats* and the most voted party did not obtain a majority of seats and ii) only includes the second and third most voted parties. Condition (ii) is similar as before and condition (i) guarantees that we focus on relevant cases, since if the most voted obtained a majority, it will appoint the mayor almost surely.<sup>31</sup>

We can thus repeat our RDD, with the running variable also redefined accordingly (the difference between second and third most voted). Figure 2a provides a graphical exercise similar to Figure 1a but, since it is based on this newly defined sample, all parties left of the cutoff are the third most voted, while all parties right of the cutoff are the second most voted. A discontinuity at the cutoff is also visible, although it is smaller and noisier than the one in Figure 1a. Note that no discontinuity is visible in the placebo graph (Figure 2b) which plots lagged outcomes—close second and third most voted parties are equally likely to have appointed the incumbent mayor. Figure 2b also increases confidence that the jump observed in Figure 2a is not driven by noise.

<sup>30</sup>The available data on deputy mayors list their party affiliation, but no their rank. Hence we cannot focus on the effect of appointing, for example, the first deputy mayor that takes office when the mayor is absent. Data on deputy mayors is not available for the entire sample, and hence sample sizes in Table A4 are smaller than in Table 1.

<sup>31</sup>Our sample contains 1,565 elections (from 1,204 unique municipalities) satisfying these two conditions.



Table 2 provides the regression results, following the template of Table 1. Column (1) indicates that a party that barely finishes in third place but almost tying with the second place (and with the same number of seats) has a 6.7% probability of appointing the mayor. That probability is almost 16% for the party finishing in second-place in such an “almost tie.” This implies that the “second most voted” label also generates an advantage compared to the “third most voted” label. Columns (2)-(4) indicate that effects of similar magnitude are estimated using different specifications. All estimates are significant at the 5% level. Panels B and C and Figure A8 provide the evidence of covariate balance (as expected in a RDD). Figure A1 shows the number of observations in each bin of Figure 2a. Figure A4 presents the robustness of the estimates to bandwidth choice.

This effect suggests that the norm generalizes to lower ranks. While of interest in itself, this effect has an important implication to the interpretation of the effect of being most voted. Since there is no status quo rule benefitting second versus third placed parties, it is impossible that such status quo rule drives the effects on Figure 2a and Table 2.<sup>32</sup> Moreover, it suggests the similar effects of being most voted (Figure 1a and Table 1) are also not driven by the status quo rule. Section 6 further discusses this issue.

**Third versus Fourth Effects.** We also estimate a similarly defined third (versus fourth) most voted effect. We redefine the sample accordingly, and Figure A9 presents a graphical analysis similar to Figures 1a and 2a. Note the y-axis range matches that of Figure 2a for comparison. While no jump at the cutoff is visible, it should be noted that such lower ranked parties rarely appoint the mayor, so proportionally similar effects would not be detectable given the lack of statistical power.<sup>33</sup>

### 3.4 Interpreting Magnitudes and Comparison to Effects of One Additional Seat

**Interpreting the Results as Norms.** We return to the effect of being most voted (instead of second most voted) discussed in Subsection 3.2 and reported on Figure 1a and Table 1. First, we highlight the interpretation of the result that was briefly discussed in the introduction. Most voted parties that obtained (almost) the same number of votes as the second most voted, should be, on average, similar in observable and unobservable characteristics (Lee 2008). Moreover, our sample conditions on both first and second most voted parties having the same number of seats.

Hence, theories of legislative bargaining and coalition formation predict the effect we estimate to be zero. This applies to both cooperative theories of coalition formation (e.g., the core, stable set, or bargaining set) and to noncooperative theories where the structure of proposals and agreements is defined. Such theories take the number of *seats* each party has as the primitives of the model, and hence do not allow a role for the vote shares conditional on the allocation of seats they determined. Allocations of seats are also the starting point on the empirical work on government formation and its allocations. The stylized fact known as Gamsons’ Law proposes that coalition governments distribute portfolios in proportion to each member party’s contribution of *seats* to the coalition. Structural models of government formation also take seat allocations as their starting point. Indeed, standard

<sup>32</sup>Spanish electoral rules do not create any differential treatment of second instead of third most voted parties.

<sup>33</sup>Our sample covers 996 elections in which the third and fourth most-voted parties tied in seats, and the third (fourth) appointed the mayor in only 25 (11) of those.



datasets used in this literature, such as the *European Representative Democracy Data Archive*, collect only information on the number of seats (and not general election votes) each party in a parliament obtained.<sup>34</sup>

This captures the key aspect to how we identify a norm. When comparing two groups that should have, given formal rules, the same bargaining power, finding differences in bargaining outcomes isolate the effect of the norm. By “norm,” we mean an informal rule or principle that binds agents and regulates proper or acceptable behavior. We highlight the importance of the *informal* aspect. Our results are best explained as the outcome of parties following a rule that is not formally coded in laws and regulations, but informally agreed upon by (part of ) members of a society. Moreover, the results suggest that parties follow the norm even when it is costly, as the example of left-wing combined majorities (the PSOE and IU) allowing right-wing mayors (from the PP). Section 4 provides a formal model where different norms can be interpreted as different equilibria.

Note also the magnitude of the effect of the norm has a nuanced interpretation. An effect of being most voted of 20 p.p. is consistent with 20% of the municipalities in our sample having a norm that always binds (the most voted party always appoint the mayor). However, it is also consistent with all municipalities in the sample having a norm that only binds with probability equal to 20%. Of course a combination of these two extreme cases is also possible. Given our research design, it is not possible to pin down which case better describes the data.

**Comparison effects of one additional seat.** While, given the discussion above, any non-zero effect of “being most voted” is perhaps surprising, it is also useful to gauge the magnitude of our effects to that of being awarded one additional seat.

A similar RDD approach can be used to estimate the effect of one additional seat on the probability of appointing the mayor. So far, we have restricted our sample to cases where the first and second most voted parties have the same number of seats. However, there are also cases where the first and second most voted almost tie in votes, but the most voted is awarded one more seat than the second-placed. Whether one additional vote leads to one additional seat is defined by the rounding inherent to D’Hondt rule.<sup>35</sup> Hence, we can estimate the effect of receiving one additional seat by restricting the sample to cases where the first and second most voted parties have different number of seats.<sup>36</sup>

Moreover, it is useful to further separate the cases where the most voted party obtains one more seat than the second most voted in three different categories.

1. The additional seat creates more *nominal*, but no more *real* bargaining power. An example is a 5-4-2 seat distribution in an 11-seat council. While the most voted party has more seats than others, it has no “real” advantage since its ability in forming coalitions is no different from the

---

<sup>34</sup>See references in footnotes 4 and 5.

<sup>35</sup>For example, the introduction describes the example of a 11-seat legislature where parties vote shares were 42.0%, 41.9%, and 16.1%. D’Hondt rule generates a 5-5-1 seat allocation for these vote shares. However, if the vote shares are the slightly different 41.0%, 40.9%, and 18.1%, D’Hondt rule generates a 5-4-2 seat allocation.

<sup>36</sup>Note that we condition the sample definition to a variable defined at the municipality-year level, which does not vary across parties within a municipality-year. This maintains the RDD “internal validity” (covariate balance) in the subsamples, as discussed in Subsection 3.1.

other parties, as any two parties can form a majority.<sup>37</sup>

2. The additional seat can generate both more *nominal* and *real* bargaining power. An example is a 5-4-1-1 seat distribution in an 11-seat council. The most voted party has an advantage in coalition formation: it only needs to add one of the parties that obtained one seat to obtain a majority, while the second most voted needs to convince both.
3. The additional seat awards a majority of seats to the most voted party (e.g., a 6-5 seat distribution in an 11-seat council).

These three cases are directly observable and we can thus define three separate subsamples accordingly.<sup>38</sup> Figure 4 provides the regression discontinuity plot these different subsamples. The blue circles correspond to the case where the two most voted parties tie in number of seats and are thus exactly the same as in Figure 1a. The red triangles plot the cases where the most voted party has one more seat than the second most voted party, but no additional bargaining power (case 1). The effect of just being labeled the most voted is similar in magnitude to obtaining one additional seat. The corresponding estimates are provided on Panel B of Table A6, which show the estimated effect is larger, but not substantially so.

However, one additional seat that is associated with more real bargaining power does provide substantially more ability to appoint mayors. These are depicted in green squares, which focus on the elections matching case 2 above. As expected, receiving a majority of seats (yellow diamonds) makes a party almost surely appoint the mayor. Panels C and D of Table A6 provide the corresponding estimates.

Overall, the effect of the most voted norm is substantial when compared to the effects of an additional seat. Based on our baseline specification (column 1 of Table A6), the effect of simply being labeled the most voted equals 60% of the effect of being labeled the most voted and obtained one more seat (but no more “real bargaining” power) and 28% of the effect of one more seat that awards real bargaining power. This suggests that the norm we study has effects of comparable size to other aspects that have been the previous focus on the literature on legislative bargaining. Previous work has noted that nominal bargaining affecting outcomes conditional on real bargaining constitutes a puzzle (Warwick and Druckman 2001, Frechette, Kagel, and Morelli 2005). The existence of the norm we study adds an additional mechanism that may help explain it.<sup>39</sup>

---

<sup>37</sup>Councils with this structure are labeled as “top-three” in Laver and Benoit (2015)’s classification of legislatures. Note also these are cases where the Shapley value or minimum integer weight of all the top three parties is the same.

<sup>38</sup>Whether a council falls in case (1) or (2) is determined by whether the second and third placed parties together form a majority or not (Laver and Benoit 2015). Of the 37,122 elections in our sample, 7.8% have the two most voted tying in seats (the focus of Subsection 3.2), and 7.9%, 8.6%, and 75.7% in cases (1), (2) and (3), respectively.

<sup>39</sup>A caveat when comparing the effects in different panels of Table A6 is that municipalities that enter each subsample may differ systematically. For example, elections that constitute “case 3” are likely to have fourth placed parties obtaining larger vote shares than those that constitute “case 2.” However, we note that the difference between one additional vote leading to one additional seat is mostly driven idiosyncratic rounding that is inherent to D’Hondt rule, as exemplified in footnote 35.

## 4 Theoretical Framework

The model presented in this section is stylized and abstracts from some aspects of government formation and other mechanisms that possibly play a role in explaining our results. However, it illustrates a specific mechanism and yields predictions that help guide the subsequent discussion of the mechanisms driving our results.

Our starting point is a canonical framework of political accountability (Barro 1973, Ferejohn 1986, Persson, Roland, and Tabellini 1997). We add to it not only legislative bargaining but also a role for elections in aggregating diffuse information. *After* an election, voters can infer information about an uncertain state of the world from vote shares. This informs voters on which party they prefer would appoint the mayor. However, parties' representation in a council is set at this point and they may bargain and form coalitions based entirely on rent allocations, ignoring voters' interests. This creates a conflict of interest between voters and parties and a reason for the former to discipline the latter. The model has multiple equilibria, which can be interpreted as norms that voters may adopt.

**Setup.** A large (odd) number of identical and infinitely lived voters maximize  $E \sum_{t=0}^{\infty} \delta^t u_t$ , where  $0 < \delta < 1$ ,  $E$  is the expectations operator, and  $u_t$  is their utility. Every period, one state  $s_t$  of the world is realized. There are three possible states:  $s_t \in \{A, B, C\}$ . There are also three types of parties ( $A$ ,  $B$ , and  $C$ ), of which one must appoint the mayor. Voters receive positive utility if the mayor's type matches the state of the world.  $u_t = 1$  if  $m_t = s_t$  and  $u_t = 0$  if  $m_t \neq s_t$ , where  $m_t$  denotes the party of the mayor. This can be interpreted as different possible events occurring, each of them being better dealt with by a specific type of party, or only one party in each period having a competent leader, and which one being uncertain.

Each party also maximizes an expected utility function,  $E \sum_{t=0}^{\infty} \delta^t x_t$ , where  $x_t$  denote the rents they obtain from office:  $x_t = 1$  if the party appoints the mayor, and zero otherwise (i.e., rents are indivisible).<sup>40</sup> Bargaining follows a specific procedure. If one party received a majority of votes in the previous election, it can unilaterally choose which party appoints the mayor. If no party had a majority of votes, then one party is randomly "recognized" (i.e., selected to propose which party appoints the mayor). All parties then vote on whether to accept or not this proposal. If one of the two non-recognized parties accepts, the mayoral appointment is realized. If not, party  $A$  appoints the mayor. This procedure thus matches the one round of voting by majority rule feature of Spanish municipalities.<sup>41</sup> The choice of party  $A$  as the status quo is without loss of generality and made to illustrate how status-quo rules play no role in our argument. We do not specify parties' recognition probabilities, assuming only they are a continuous function of previous election's vote shares.<sup>42</sup>

Parties' preferences and the bargaining procedure are thus independent of voters' welfare and the

---

<sup>40</sup>This can be understood as the mayor setting a specific policy to the preferences of the party, or mayors not being able to commit to share the spoils of office with its supporting coalition. Appendix B discusses how this can be relaxed.

<sup>41</sup>This structure is analogous to assuming parties have votes weighted by their previous election's votes and decide by majority rule. With three parties, if one does not obtain a majority of seats, then any two parties have a combined majority. This matches the three-person majority game structure of municipalities in our sample discussed in Section 2.

<sup>42</sup>For example, recognition probabilities being  $1/3$  for all parties or the same as the vote share in the previous election satisfy this condition. Continuity rules out the most voted party being recognized with certainty, which would make the model unattractive to study our empirical results.

states of the world. This creates, in a stark but tractable manner, a dissonance between voters' and parties' interests that is a key feature of the model. Voters prefer the mayor that matches the state to be appointed, but the choice of mayor may be determined by factors that are orthogonal to their interest. If states were directly observable, voters could easily address this issue by awarding a majority to the party matching the state.

**Uncertainty and information structure.** However, states of the world are never directly observed by voters or parties. The probability that state  $s$  occurs in period  $t$  is denoted  $p_t^s$ , with  $\mathbf{p}_t$  denoting the vector  $[p_t^A, p_t^B, p_t^C]$ . Moreover, voters and parties face uncertainty about the vector  $\mathbf{p}_t$ , which is drawn every period from a (common knowledge) distribution  $G(\mathbf{p})$  that is serially uncorrelated and identically distributed over time. Each voter individually observes a signal  $\sigma_t$  about the state of world every period. The three possible signals are also  $\{A, B, C\}$ , with the probability the signal is  $s_t$  given by  $p_t^s$ , drawn independently for each voter.<sup>43</sup>

Hence, each period a voter updates her beliefs about the state of the world twice. At the start of the period, all voters have the same priors based on the expected value of  $G(\mathbf{p}_t)$ . After she observes her private signal of value  $i$ , she forms a new belief  $\Pr(s_t = k | \sigma_t = i)$  for all  $k \in \{A, B, C\}$ , which informs her vote decision. Finally, after observing the election results, she updates again, based on other voters' strategies and election results. If all citizens vote according to their signals (e.g., vote for party of type  $A$  if signal is  $A$ ), then she will expect the probability that the state is  $s$  to be the vote share of party of type  $s$ . However, by the time this information is revealed, parties' representation in the legislature is already defined and, by the time another election occurs, a new vector  $\mathbf{p}_{t+1}$  and state  $s_{t+1}$  will be drawn, making previous information irrelevant.

We assume that  $G(\mathbf{p}_t)$  is such that  $\Pr(s_t = i | \sigma_t = i) > \Pr(s_t = j | \sigma_t = i)$  for all  $i \neq j$ . This implies that, after a voter observes a private signal of value  $i$  (but before observing election results) she expects  $i$  to be the most likely state and prefers party of type  $i$  to appoint the mayor. Appendix A provides an example of a  $G(\mathbf{p}_t)$  function and illustrates how voters update in the model.

**Timing and elections.** The sequence of events is the following. At the start of every period  $t$ , nature draws the vector  $\mathbf{p}_t$ . Based on this vector it draws the state of the world and the signals each voter observes. Each voter then chooses how to cast a vote. There are six possible votes to cast: voting for one of the parties  $A$ ,  $B$ , or  $C$  that were in office in the preceding period (the incumbents) or voting for one of challenger parties  $A$ ,  $B$ , or  $C$ . In other words, for each of the three types of party, there is always a challenger party of the same type that is identical in all respects to the incumbent. An incumbent that receives zero votes is never re-elected again. Parties then appoint the mayor according to the procedure described above. Payoffs are realized and a new identical period starts over.

**Discussion of assumptions.** This setup captures a dual role for elections. They can serve as an information aggregation mechanism and also as a way to discipline incumbents to behave in consonance with voters' interests. The assumption that there is always a challenger party of each type makes this dual role clearer. While alternative assumptions that would lead the choice to punish one incumbent also reward the other incumbent parties would perhaps be more realistic, they would complicate the model and create a conflict between the information aggregation and disciplining roles

<sup>43</sup> All realizations of  $G(\mathbf{p})$  satisfy  $p^A + p^B + p^C = 1$ .

of elections. Note, however, that a large number of parties, with presence at the national, regional, and municipal level, operate in Spain. If the types of parties are interpreted as their ideologies (e.g., left, center, and right), this can be interpreted as, say, multiple leftist parties (some regional or municipal) that can replace one another.<sup>44</sup> While the model assumes a specific bargaining procedure, Appendix B discusses how the results are robust to assuming multiple rounds of bargaining and allowing rents to be divisible across parties.

**Equilibria.** We restrict our attention to sequentially rational equilibria in which every voter chooses a pure strategy that conditions her decision on her last observed signal and the incumbent party’s behavior in the preceding period. All parties choose pure strategies that condition only on the result of that period’s election.<sup>45</sup> We highlight that the information structure is serially uncorrelated: a new independent draw of  $\mathbf{p}_t$  and the state of the world is drawn every period. Hence, events from past periods carry no information about the current state of the world.

This model has multiple equilibria, similarly to Barro (1973), Ferejohn (1986), and Persson, Roland, and Tabellini (1997). Since an incumbent party is identical to a challenger of the same type, voters find choosing either an incumbent or challenger (of the same type) ex post optimal. Moreover, voters conditioning their choices on incumbent’s previous behavior is also optimal. Different equilibria where voters condition their choices on incumbent behavior or not, or condition in different ways, can be interpreted as different norms: simple conventions voters can adopt. Since they are equilibria, they are also self-enforcing (given everyone follows the norm, each individual also finds it optimal to do so too). This interpretation of multiple equilibria as different norms in a model with similar structure is discussed in Persson, Roland, and Tabellini (1997).

We do not fully characterize the equilibria in this model, but focus on two cases: one equilibrium with the “most voted appoints the mayor” norm and one without it. We begin with the latter.

**Proposition 1.** There exists an equilibrium where, every period, a citizen observing signal  $\sigma_t = i$  votes for the incumbent party of type  $i$ . A party that obtains a majority of votes appoints the mayor. If no party obtains a majority, then each party, if recognized, makes an offer to appoint the mayor itself. All parties, if not recognized, accept any proposal.<sup>46</sup>

In this equilibrium, if no party receives a majority of votes, each party has a chance of appointing the mayor equal to their recognition probability—which must be the same for two parties that tied in votes. Hence, this equilibrium does *not* generate a “jump” in the RDD studied in Section 3. Those results, however, can be captured by the following equilibrium, where the most voted party appoints the mayor in every period.

---

<sup>44</sup>We restrict our attention to equilibria where all voters receiving the same signal cast the same vote and where two parties of the same type will not receive votes in an election. We hence abstract from the possibility of two parties of the same type being represented to keep the exposition concise.

<sup>45</sup>The restriction on pure strategies is realistic for a large number of voters, as it would be difficult to coordinate on a strategy that is random from the viewpoint of the parties. The restriction to strategies that are not dependent on longer histories is made for ease of exposition.

<sup>46</sup>*Proof:* Voters’ strategies being a best response was discussed previously. Since all offers are approved in equilibrium, all parties find it optimal to offer to appoint the mayor themselves in every period. Since all parties accept any proposal, any deviation by an individual party cannot change the bargaining outcome (since a majority is accepting). Note this self-fulfilling feature is not needed for this result. A similar equilibrium can be sustained with party  $A$  rejecting any proposal that does not make it the mayor. Parties  $B$  and  $C$  will still be indifferent between accepting or not any offer that does not appoint them the mayor (since  $A$  is the status quo).

**Proposition 2.** If  $G(\mathbf{p}_t)$  is such that three conditions are satisfied: i)  $\Pr[p_t^A > \max(p_t^B, p_t^C)] > 1 - \delta$ ; ii)  $\Pr[p_t^B > \max(p_t^A, p_t^C)] > 1 - \delta$ , and iii)  $\Pr[p_t^C > \max(p_t^A, p_t^B)] > 1 - \delta$ , then there exists an equilibrium where, every period, a citizen observing signal  $\sigma_t = i$  votes for the challenger of type  $i$  if, in the previous period,  $i$  both appointed the mayor and was not the most voted party. If, in the previous period,  $i$  did not appoint the mayor or did so after being most the voted, a citizen observing signal  $\sigma_t = i$  votes for the incumbent of type  $i$ . All parties, if recognized, propose that the most voted party appoints the mayor. The most voted party accepts a proposal in which it appoints the mayor, but rejects all other proposals. The second (third) most voted party rejects a proposal in which it appoints the mayor, but accepts all other proposals.<sup>47</sup>

The key intuition behind Proposition 2 is that a second or third most voted that is recognized compares the utility of appointing the mayor for one period and never being reelected again with the continuation value of being reelected. The latter is the perpetuity of the probability of being the most voted party, which conditions (i)-(iii) guarantee is smaller than the one-period gain from deviating from the norm.

**Interpretation of the norm.** While the “most voted appoints the mayor” norm is associated with an equilibrium where agents strategically play best responses, it can also be interpreted as players following a simple heuristic or rule-of-thumb. Voters reelect the party that they perceive as the best one for future conditions, but punish at the polls a party that appointed the mayor but was not the most voted. This norm can be enforced simply by the notion that it is “unfair” or “undemocratic” for a party that did not win the most votes to appoint the mayor. Parties expect this and know they will be punished if they deviate from the norm.

While this norm is simple, in this particular model it maximizes expected voter welfare. It guarantees the party most likely to match the state of the world appoints the mayor in every period. This is not the case in the equilibrium described in Proposition 1. Moreover, given that players do not directly observe the states of the world, it maximizes voters’ expected utility. Our theoretical framework does not directly address equilibrium selection: why the the norm is adopted or not. However, the fact it is optimal for voters can provide an explanation for its prevalence.<sup>48</sup>

Note that the equilibrium in Proposition 2 generates the RDD “jump” we study in Section 3. Even when the two most voted parties are only one vote apart, the most voted appoints the mayor. This occurs even though voters (rationally) understand that the difference in expected welfare between appointing the first and second most voted is close to zero. This highlights the interpretation of

---

<sup>47</sup>*Proof:* Voters’ strategies being a best response was discussed previously. The strategy of the most voted party is clearly a best response. For the second and third most voted, a deviation to proposing appointing the mayor itself will be accepted and lead to a payoff of one in that period. However, it will trigger the voters’ punishment strategy and make it never be elected again. The value of not deviating and continuing to be re-elected is thus the perpetuity of the probability of being the most voted party (e.g.,  $\frac{\Pr[p_t^A > \max(p_t^B, p_t^C)]}{(1-\delta)}$  for party  $A$ ), which is larger than one for all parties given conditions (i)-(iii). Hence, proposing the most voted party appoints the mayor is a best response. Strategies are also best responses regarding accepting offers: the second most voted party is indifferent between accepting or not a proposal that makes the first or third most voted party the mayor. A similar logic applies to the third most voted.

<sup>48</sup>Proposition 2, however, specifies the necessary conditions for the equilibrium with the norm we study, and hence clarifies partly why it might not occur in some contexts. Note that it is possible to construct an equilibrium where voters enforce a norm that the second (or third) most voted must appoint the mayor. However, voters’ expected welfare would be lower in such equilibrium.



equilibria as norms: the behavior of voters and parties is mutually self-enforcing. We also believe the interpretation of equilibrium behavior as a heuristic or rule of thumb is useful. Voters demanding that the most voted party appoints the mayor is consequential to welfare in the cases the most voted party has substantially higher vote share than the second most voted. However, this behavior becomes coded as a heuristic based on ranks, which is applied even in the cases where the consequences are minimal (parties almost tying). A similar interpretation is that ranks are salient but the continuous variable that determines ranks are not, as suggested by the behavior of political agents (Anagol and Fujiwara 2016; Folke, Persson, and Rickne 2015), consumers (Pope 2009), and investors (Hartzmark 2015).<sup>49</sup>

**Empirical Implications.** In the equilibrium described in Proposition 2, the most voted party always appoints the mayor. The results in the previous section, however, indicate that the second placed parties appoint the mayor with non-trivial frequency. There are two possibilities to reconcile this fact with the model. The first possibility is that not all municipalities in the sample are in the equilibrium with the norm. Some are in an equilibrium in which the two parties tying in seats have equal probability of appointing the mayor. Under this interpretation, the effects from Section 3.2 pin down the share of municipalities that follow the norm.

Another possibility is to incorporate (exogenous) deviations from the norm in the model. These can occur due to “trembling hand shocks” in parties’ proposal strategies that create the possibility of second and third placed parties appointing the mayor on the equilibrium path. This second possibility has the advantage that it predicts that voter punishment for such second and third placed parties can be observed. This motivates the evidence below suggesting voters punish second most voted parties that deviate from the norm.

The structure of the model is such that voters also prefer the second most voted party to appoint the mayor instead of the third most voted party. Hence, the model could be extended to generate a norm that awards an advantage to the second most voted party over the third most voted. Such extension can be derived from an exogenous impediment to the most voted party appointing the mayor (e.g., a fixed probability the most voted party is not allowed to appoint the mayor in a period). In such cases, norms where voters impose that the second most voted take preference over the third most voted in appointing the mayor can also be an equilibrium.

Finally, the model makes a prediction of when an equilibrium with “most voted party appoints the mayor” norm can occur. The conditions for this equilibrium to exist are such that, for any party, the benefit of deviating from the norm is smaller than the benefit of following the norm, so it can obtain rents in future periods when it is the most voted party. Whether this condition is satisfied is not directly observable in the data. However, it can be approximated since the condition is more likely to be met when the third most voted party has a larger share of the votes. Intuitively, if the vote shares of parties are 45.5%, 44.5%, and 10%, the third most voted party is unlikely to be the first most voted in future elections: it has to close a 35.5 p.p. vote share difference. Compare that to an election where the vote shares are 35.5%, 34.5%, and 30%. It is more likely that the third-placed party will be

---

<sup>49</sup>As discussed in the introduction, the salience of ranks can also be understood as a case of “limited attention” in which a decision maker simplifies decisions by processing only a subset of available information.



the most voted in subsequent elections. This suggests an additional test based on the heterogeneity of effects: the “effect of being most voted” should be stronger when the third most voted party has higher vote shares.<sup>50</sup>

## 5 Empirical Implications of the Theoretical Framework

### 5.1 Do Voters Punish Second-Placed Parties That Break the Norm?

This subsection provides evidence suggesting that voters enforce the norm we study. First, it should be noted that voters appear to explicitly agree with the norm. For example, a nationally representative poll found that 55% of Spaniards agree that “*it is more democratic that the most voted party forms the government, even if that party does not have an absolute majority of the votes*” (El Pais 2015). Second, parties and candidates also incorporate the norm into their political discourse. Leaders of both major national parties (PP and PSOE) have made campaign promises to not form government if their party was not the most voted in both national and local elections (Europa Press 2007).<sup>51</sup>

Second, we estimate the effect of appointing the mayor on future vote shares. In particular, we test whether second most voted parties that appoint the mayor go on to lose votes, compared to first most voted parties that appoint the mayor. For identification purposes, we focus on the cases where the first and second place parties have a vote share difference of 1 p.p. or less and the parties tied in seats (i.e., the cases close to the cutoff in the RDD analysis described in Section 3).<sup>52</sup> However, there is still the issue that parties select into whether or not they appoint the mayor, which we address by analyzing pre-existing trends later.

We estimate the following triple-difference equation for vote share ( $v$ ) of party  $p$  in municipality  $m$  at year  $t$ :

$$(v_{pm,t+1} - v_{pm,t}) = \alpha + \beta m_{pmt} * f_{pmt} + \gamma m_{pmt} + \delta f_{pmt} + \epsilon_{pmt}, \quad (2)$$

where  $m = 1$  if party  $p$  appointed mayor,  $f = 1$  if party is first-place. Given that the sample includes only first or second most voted parties,  $\gamma$  is the effect of appointing mayor for the second most voted party and  $\gamma + \beta$  is the effect of appointing mayor for the first most voted party. Our hypothesis is that  $\beta > 0$ : first most voted that appoint mayors are rewarded compared to a second most voted party that also appoints a mayor. Given the first-differences specification, equation (2) nets out municipality-party fixed effects and time effects are absorbed into  $\alpha$ . Standard errors are clustered at the municipality level.

The results are shown in Column (1) of Table 3, which shows that first placed parties that appoint mayors observe a subsequent growth in vote shares that is 4.8 p.p. larger than second placed parties appoint mayors. Columns (2) and (3) show similar results when a full set of province and party in-

---

<sup>50</sup>Since our results are identified from the cases where the two most voted parties almost tie, a larger vote share of the third most voted is associated with lower vote share of the second and first. Moreover, it is not possible to study heterogeneity by the vote share of the first most voted, holding the second most voted party’s vote share constant.

<sup>51</sup>Similar pledges have been made at regional governments, for example in Andalusia (ABC 2004). Section 6 discusses similar cases in other countries.

<sup>52</sup>Sample size is smaller than the one used on column (2) of Table 1 since not all parties run in two consecutive elections, or local parties change names making it impossible to identify them over time.

dicators is added (which control for province-time and party-time variation, given the first-difference specification). As a placebo test and a check of whether first and second placed parties are on “parallel trends,” column (6) estimates equation (2) with lagged outcomes. It finds no significant effect, suggesting the previous result is not driven by a pre-existing trend.

To further probe the dynamics of the effect, Figure 5b provides the event-study counterpart for equation (2). In particular, we estimate the following equation:

$$(v_{pm,t+k} - v_{pm,t}) = \alpha_k + \beta_k m_{pmt} * f_{pmt} + \gamma_k m_{pmt} + \delta_k f_{pmt} + \epsilon_{pmt} \quad (3)$$

separately for  $k$  equal to -3, -2, -1, 1, 2, and 3. Figure 5b plots the  $\beta_k$  against  $k$ , as well as their 95% confidence intervals. The graph indicates no pre-existing trends (i.e., zero placebo effects on lagged outcomes) and suggests that the differential effect of a mayoral appointment for first-placed parties dissipates after two elections, although perhaps not fully.

Given the triple-difference nature of the estimation, it is not clear whether the effects of Figure 5b are driven by most voted parties gaining more votes than second most voted parties that do so, or the latter losing votes. In other words, the effects are relative to the counterfactual of the other party (and can thus be interpreted as a “reward” for the most voted or a “punishment” for second most voted when they appoint a mayor). To illustrate this issue, Figure 5a provides the double-difference event study graph for both second placed parties and first placed parties separately. In particular we estimate equation 3 separately for only second most voted parties (red squares) and first most voted parties (blue circles).<sup>53</sup> While the second most voted party that appoints a mayor gains votes (over a second most voted party that does not), this can to be explained by the continuation of a pre-existing trend (parties that appoint mayor are in positive trajectories). Remarkably, first most voted parties are on a similar trajectory before appointing a mayor, but go on to gain even more than second placed parties after their appointment.

Finally, the model also suggests that third most voted parties would also be punished if they appointed the mayor. However, given the focus on the two most voted parties tying in seats, providing a similar triple-difference specification for third most voted parties is not possible.

## 5.2 Heterogenous Effects by Strength of the Third-Placed Party

Another prediction of the model is that, in elections where the third most voted party has a higher vote share, the norm we study is more likely to occur, and hence the effect of being most voted should be stronger. Table 4 provides evidence supporting this prediction. It repeats our estimation of the main results (Table 1), but separating the sample into the cases where the third most voted party vote share is above (Panel A) and below (Panel B) the median.<sup>54</sup> The effects are substantially larger in Panel A, and it is possible to reject that the effects in both subsamples are the same in all

<sup>53</sup>This implies we estimate  $(v_{pm,t+k} - v_{pm,t}) = \alpha_k + \beta_k m_{pmt} + \epsilon_{pmt}$  when using only the second most voted parties and  $(v_{pm,t+k} - v_{pm,t}) = (\alpha_k + \delta_k) + (\beta_k + \gamma_k) m_{pmt} + \epsilon_{pmt}$  when using only the first most voted parties.

<sup>54</sup>We use the median of the sample with optimal bandwidth defined in Subsection 3.2 (Table 1), which is a vote share of 16.5%. In the above (below) median subsample, the average vote share of the top two most voted parties is 40% (33%) each, with the third most voted obtaining 12% (22%).

specifications across columns (at the 5% level). Figure A10 provides the graphical counterpart.<sup>55</sup>

Moreover, columns (4) and (5) of Table 3 estimate equation (2) separately for each subsample. It thus tests whether the voters’ punishment for deviation of the norm is more evident in elections where the third-placed party is stronger. We find that this is indeed the case, although the estimates are noisily estimated (likely given the smaller subsamples).

## 6 Alternative Explanations

**Status-quo rule.** To the best of our knowledge, the only differential institutional treatment of parties by rank of their votes in Spanish municipal elections is the status quo described in Section 2. If no candidate receives a majority of votes in the council election, the party with the most votes appoint the mayor. While, at first pass, this appears to be likely to explain our results, there are four reasons we believe the status-quo rule cannot be the main driver of our results.

First, and perhaps most importantly, there is no similar status quo rule, or any other institutional advantage, that is given to the second most voted party. Hence, the status quo rule cannot play a role in explaining the second most voted versus third most voted effects described in Section 3.3 and thus cannot account for the entirety of our evidence. Second, it is not clear why the existence of this status quo would make voters “punish” second most voted parties that deviate from the norm or why it would interact with the vote share of the third most voted party (Section 4) or the size of municipality or council (Figure A3).

Third, note that parties’ coordination failures or mistakes when casting votes for mayor are unlikely to generate our effects via the status-quo rule. As noted in Section 2, a majority of council members can easily replace the mayor at their will at any point of the term. Hence, even if by mistake in casting votes a majority was not obtained and the first-placed party appointed a mayor that displeased a majority, that could be undone quickly.

Fourth, also as discussed in Section 2, the situation of the vast majority of councils in our sample fits a three-player majority game: any two of the three most voted parties can form a majority coalition. In such cases it is particularly unclear why the status quo should matter. To formalize this argument, Appendix C outlines a voting game that approximates these conditions and the Spanish institutions for selecting a mayor. It shows that, given sensible equilibrium refinements (i.e., parties not taking weakly dominated actions or allowing two parties to coordinate in their deviation), the status-quo rule is irrelevant in defining which party elects the mayor.

**Agreement among parties.** Another possible explanation for our results is that parties create a (perhaps implicit) agreement that the most voted party should form the government. This is a distinct mechanism from the one suggested by the theoretical framework only if the reason for such agreement is not that voters would enforce the norm. In other words, the theoretical framework shows how a

---

<sup>55</sup>Figure A10 is constructed similarly to Figure 1a, but for each subsample. Figure A10b illustrates why the effect for the below median subsample varies across columns in Table 4, there is nonlinearity close to the cutoff that is not captured by the specifications using the entire sample, which find a larger effect than local estimates (columns 1-2). Figure A3 shows that this heterogeneity holds within some council sizes (e.g., the effect appears larger in 3-3-3 councils than 4-4-1 councils, or 4-4-3 instead 5-5-1).

norm that arises from voters’ strategies determine parties’ behavior.

It is not straightforward why causality would run in the other direction. Even if parties decided on their own to enforce the “most voted party appoints the mayor” norm, it is not clear why voters would punish a party that deviates from it. Similarly, it is not clear why such agreements would be more common when the third placed party obtained more votes. Additionally, it is not evident why parties would find this norm desirable. One possibility is that if bargaining after every election is costly, the norm would be in their interest. However, the costs of bargaining seem small compared to the importance of a mayoral appointment. If that is the case, there would be strong incentives for second or third most voted parties to systematically renege on this agreement, eventually making its effect disappear.

One, albeit indirect, test of this mechanism is that the effect of being the most voted should be stronger in municipalities with more frequent cases of ties in number of seats.<sup>56</sup> Panel A of Figure A11 repeats our main RDD plot (Figure 1a) separating the sample into cases from municipalities that experienced multiple cases of the two most voted parties tying in seats, and those that only experienced one case, during our sample period.<sup>57</sup> The effects are similar in both subsamples. The same applies when looking at cases with even more frequent ties (two or more) in Panel B. There is no evidence that municipalities where ties in seats occur frequently are more likely to present the norm we study.

## 7 Evidence from Other Contexts

The main reason we focus on Spanish municipalities, is that, given its large number of municipalities (with small council sizes), we can observe a large number (2,898) of elections where the two most voted parties tie in seats. To the best of our knowledge, there is no other context with available data and a comparably large number of “ties in seats.”<sup>58</sup>

However, a natural question is to what extent our results are specific to Spain. In this section we briefly discuss anecdotal evidence of similar issues in American and British national politics and additional evidence from 28 European parliamentary democracies.

**Anecdotal evidence from the USA and UK.** One aspect of our norm is that it may transform a majoritarian procedure (parties representing a majority support a mayor) into a pluralitarian procedure (the most voted party appoints the mayor).<sup>59</sup> A similar issue surrounded the 2016 presidential primaries in the US. The procedures of the Republican Party make clear that the vote of a *majority* of delegates is needed for a nomination. However, before early May, many expected Donald Trump to achieve a plurality, but not a majority, of delegate votes. Commentators (and Trump himself) declared that the candidate with the most delegate votes should be the nominee. More interestingly, 62% of

---

<sup>56</sup>Presumably, previous experience with costly bargaining would be the reason for norm to arise in future cases.

<sup>57</sup>Note that a municipality must experience at least one case to enter the sample. This definition separates the sample into two subsamples of approximately same size.

<sup>58</sup>Note contexts where the executive is directly elected (e.g., Brazilian or Italian mayors) are not suitable for our analysis. Other countries with parliamentary local governments have a smaller number of municipalities, making it unlikely ties are commonly observed (e.g., Finland has 311 municipalities and Sweden has 290, while our Spanish sample is based on over 5,900 municipalities).

<sup>59</sup>See, for example, our discussion of the 2011 Olivenza election in Section 3.2.

Republican voters agreed with the statement that with “no delegate majority, the GOP should be the one with the most votes.”<sup>60</sup> Moreover, Silver (2016a, 2016b) argued that one reason for Trump’s eventual success at securing a majority of delegates was that “*Republican voters were swayed by Trump’s arguments that the candidate with the most votes and delegates should be the nominee.*”

Similarly, during campaigning for the 2015 British general election, the leader of the Liberal Democrat Party (Nicholas Clegg) stated that “*the party that gets the most votes and most seats, in other words the party that gets the biggest mandate from the British people, even if it does not get a slam-dunk majority, it seems to me right to give that party the space and the time to try and settle a government*” (Perraudin 2015).

**Evidence from European National Parliaments.** To study the possible role of a similar norm in other contexts, we study government formation in 28 European parliamentary democracies in the 1944-2010 period.<sup>61</sup> The data contains the number of seats of each party in the lower house of national legislatures, as well as the party affiliation of the appointed executive (e.g., prime minister). The countries in the sample vary both in the role such appointed executive has in government (e.g., importance of the British versus French prime ministers) and also the specific rules for government formation: “*different legislatures play different formal roles, and thus have different levels of influence, in the government formation process*” (Cheibub, Martin, and Rasch 2015).

In the vast majority of cases in our sample, there is no formal (e.g., a law or constitutional clause) institutional advantage for the party with the most seats in forming government.<sup>62</sup> It is possible that some of these countries have established informal rules (e.g. a head of government such as a monarch is expected to ask the most represented party to first formateur). However, note first that, to the extent that these are not formally coded in laws, they can be understood broadly as norms. While the specific mechanisms behind them might be different than the one modeled on Section 4, the theoretical framework might explain why the norms would exist and be supported by voters in first place. Second, Diermeier and Merlo (2004) show that there is little evidence that formateurs are chosen by rank in the elections (e.g., the first placed party is the first formateur) in a sample of National European Parliaments.<sup>63</sup>

There are few cases of the two most voted parties tying in the number of seats and we do not

---

<sup>60</sup>Note the similarity to the Spanish survey discussed in Section 5. The other option in the survey was “GOP nominee should be the best party standard-bearer,” which 33% of respondents agreed with. The survey occurred in April 2016 (Murray 2016 and Flegenheimer 2016).

<sup>61</sup> The dataset is the The European Representative Democracy Data Archive (Andersson and Ersson 2014) and the countries included are Austria, Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom.

<sup>62</sup>The exceptions are Bulgaria and Albania, which in 1993 and 1998, respectively, stipulated that the most voted party should be the first formateur (the party with power to make a proposal of government formation that is put to a vote).

<sup>63</sup>More specifically, Diermeier and Merlo (2004) show that the formateur selection is better explained by selection probabilities being proportional to seat shares (“proportional selection”) instead of ranks in seats (“selection-in-order”). Proportional selection cannot generate a jump at the RDD cutoff we present in this section. Moreover, our results are not inconsistent with Diermeier and Merlo (2004). First, they study *formateur* selection, while we study the eventual prime minister appointment. Second, their sample contains 11 European countries, as opposed to the 28 in our sample. Third, their maximum likelihood exercise estimates a model that best fits all cases in the data, while we focus on behavior near our RDD cutoff.

observe general election vote shares (only seat shares) of parties.<sup>64</sup> Hence, we cannot apply the same tests as the ones reported on Section 3, which are based on parties tying in number of *seats*. We instead estimate an analogous RDD for the effect of obtaining the most seats. More precisely, we estimate equation (1) with the running variable now defined as the *seat* share difference between the two parties with most seats in the legislature and the outcome being an indicator for whether the party appointed the prime minister. The sample contains only the two parties with the most seats in the legislature. Intuitively, we estimate the effect of obtaining one additional seat that makes a party switch from having the second most to the most number of seats in a legislature. The outcome of interest, which we label “appointing the prime minister,” is a dummy indicating whether the party appointed the relevant executive member for the entire term following the election.<sup>65</sup>

Figure 6a provides the relevant RDD plots, constructed in a similar manner to Figure 1a. The “jump” at the cutoff suggests that a party with “one more seat” than the party with the second most seats is almost 40 p.p. more likely to appoint the prime minister. Note, moreover, that the relationship between the outcome and running variable is relatively flat on the left of the cutoff. This suggests that as second-placed parties increase their number of seats (relative to the first-placed), they are not more likely to appoint the prime minister. However, the additional seat that “flips” a party into being the most represented has a sizable impact. Figure 6b provides the lagged outcome placebo and Table 5 provides the corresponding estimates. The results are overall more noisily estimated than the ones in Section 3, as there are fewer observations in the sample (Figure A12). However, they are quantitatively large and statistically significant at the 1% level.<sup>66</sup>

Differently from the results based on Spanish municipalities, the effects on Table 5 can be explained by the party with most seats having an advantage in forming coalitions. For example, it is possible that the first placed party can form a majority coalition with the third-placed party, while the second placed cannot. To investigate if such “real bargaining power” differences drive our results, as opposed to a norm of the most represented party forming government, we implement a test similar in spirit to that of Subsection 3.4. In particular, Figure A15 repeats the exercise of Figure 6a, but using a sample that excludes all cases where the party with the most seats could form a majority coalition with the third placed party, while the second could not. The effects are similar to those from the main sample. This suggests that the effect of having most seats is more likely driven by a norm than by the “real

<sup>64</sup>Our original dataset contains three cases in which the first two parties tied in seats: the Netherlands in 1952, and Belgium and Estonia in 2003. In the Dutch and Belgian cases, the most voted party appointed the prime minister. In the Estonian, the second most voted party did. These cases are excluded from our estimating sample.

<sup>65</sup>Differently from the Spanish municipal case, the timing of elections is endogenous to the support a prime minister receives. Hence it is not possible to define a similar “appointing the executive for 3/4 of the term” outcome, since a change in prime minister is usually associated with new elections taking place. Our results are similar if the outcome is defined as appointing the first prime minister after the election.

<sup>66</sup>The robustness to bandwidth choice is presented on Figure A13. We also provide a test of covariate balance based on party identity (similar in spirit to Panel C of Table 1 and Figure A5). [Andersson and Ersson \(2014\)](#) classify parties into 12 ideological families (e.g., social-democratic, liberal, communist, green). We regress our outcome (appointing prime minister) on a set of dummies indicating each family. We then use the predicted value for this regression as an outcome (Panel C of Table 5) and Figure A15. If a particular type of family was systematically more likely to be, say, just right of the cutoff, we would expect an effect on this predicted outcome. Our results suggest the opposite: parties’ ideology is balanced around the cutoff.



bargaining power” associated with more ability to form different coalitions.<sup>67</sup>

Lastly, we explore whether the effect of having most seats is driven by an increased probability of being in the ruling coalition (i.e., having cabinet positions) or an increased probability of appointing the prime minister conditional on being in the ruling coalition. Figure A16 shows the data is consistent with the latter case. It replicates the exercise of Figure 6a, but using a dummy equal to one if the party is in the ruling coalition as the outcome. We do not see a jump at the cutoff.<sup>68</sup>

## 8 Conclusion

Our main result indicates that simply being labeled the “most voted” has, in itself, a substantial effect on parties’ bargaining outcomes in a legislature. This result is difficult to reconcile with existing theories of multilateral bargaining and coalition formation. The overall evidence we provide is consistent with the existence of a norm (an informal rule) that voters enforce by punishing parties that deviate from it. This can explain why parties follow the norm even when it suggests a different course of action than natural considerations such as parties that are ideologically close forming a winning coalition.

We believe our results raise several questions for future research. The first is investigating if (and why) similar norms occur in different contexts—our results from a sample of 28 national European parliaments suggest it extends beyond Spanish municipalities. The second are the norm’s policy and welfare consequences. In particular, we highlight that the norm can turn a proportional system into one that works similarly to plurality rule—a system that is designed to allocate political power proportionally can become one that is closer to a “winner takes all” contest. This has been highlighted as a key point in how different electoral rules lead to different policies (Lizzeri and Persico 2001).

## References

- ABC (2004). Chaves acepta el reto de Teofila y promete no gobernar si no es el mas votado. [http://sevilla.abc.es/hemeroteca/historico-22-01-2004/sevilla/Andalucia/chaves-acepta-el-reto-de-teofila-y-promete-no-gobernar-si-no-es-el-mas-votado/\\_183979.html](http://sevilla.abc.es/hemeroteca/historico-22-01-2004/sevilla/Andalucia/chaves-acepta-el-reto-de-teofila-y-promete-no-gobernar-si-no-es-el-mas-votado/_183979.html).
- Agranov, M., J. K. Goeree, J. Romero, and L. Yariv (2017). What makes voters turn out: The effects of polls and beliefs. *Journal of the European Economic Association*, jvx023.
- Anagol, S. and T. Fujiwara (2016). The runner-up effect. *Journal of Political Economy* 124 (4), 927–991.
- Anagol, S. and H. H. Kim (2012). The impact of shrouded fees: Evidence from a natural experiment in the Indian mutual funds market. *The American Economic Review* 102(1), 576–593.

---

<sup>67</sup>Following Laver and Benoit (2015) classification of seat allocations in legislature, the sample in Figure A15 excludes the seat compositions with a “dominant party.” A caveat in comparing Figure A15 to Figure 6a is that parliaments that form the subsample may differ systematically. For example, countries with concentrated party structures are more likely to have a “dominant party” in the legislature. Note that in the parliaments in Figure A15’s sample, it is still possible that the party with most seats has higher “real bargaining power” than the party with the second most seats. Ideally, one would restrict the sample to cases where the top two parties have the same Shapley-Shubik value, Banzhaf index, or minimum integer weight. However, we are restricted by the relatively smaller sample size: only 127 elections provide the data used on Figure A15.

<sup>68</sup>Unfortunately, our dataset does not contain information on the *share* of cabinet positions each party obtained, and only provides information on whether the party is represented in the cabinet or not.



- Andersson, Staffan; Bergman, T. and S. Ersson (2014). The European representative democracy data archive, release 3. *Main sponsor: Riksbankens Jubileumsfond (In2007-0149:1-E)*. [www.erdda.se](http://www.erdda.se).
- Austen-Smith, D. and J. Banks (1990). Stable governments and the allocation of policy portfolios. *American Political Science Review* 84(3), 891–906.
- Bagues, M. and P. Campa (2017). Can gender quotas empower women? Evidence from a regression discontinuity design. *Unpublished Manuscript*.
- Baron, D. P. (1991). A spatial bargaining theory of government formation in parliamentary systems. *American Political Science Review* 85(1), 137–164.
- Baron, D. P. (1993). Government formation and endogenous parties. *American Political Science Review* 87(1), 34–47.
- Baron, D. P. and J. A. Ferejohn (1989). Bargaining in legislatures. *American Political Science Review* 83(04), 1181–1206.
- Barro, R. J. (1973). The control of politicians: An economic model. *Public Choice* 14(1), 19–42.
- Bendor, J., D. Diermeier, D. A. Siegel, and M. M. Ting (2011). *A behavioral theory of elections*. Princeton University Press.
- Bendor, J., D. Diermeier, and M. Ting (2003). A behavioral model of turnout. *American Political Science Review* 97(2), 261–280.
- Bisin, A., A. Lizzeri, and L. Yariv (2015). Government policy with time inconsistent voters. *The American Economic Review* 105(6), 1711–1737.
- Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and consumer choice. *Journal of Political Economy* 121(5), 803–843.
- Brown, J., T. Hossain, and J. Morgan (2010). Shrouded attributes and information suppression: Evidence from the field. *The Quarterly Journal of Economics* 125(2), 859–876.
- Callander, S. (2007). Bandwagons and momentum in sequential voting. *The Review of Economic Studies* 74(3), 653–684.
- Callander, S. and C. H. Wilson (2008). Context-dependent voting and political ambiguity. *Journal of Public Economics* 92(3), 565–581.
- Carroll, R. and G. W. Cox (2007). The logic of Gamson’s law: Pre-election coalitions and portfolio allocations. *American Journal of Political Science* 51(2), 300–313.
- Cheibub, J. A., S. Martin, and B. E. Rasch (2015). Government selection and executive powers: Constitutional design in parliamentary democracies. *West European Politics* 38(5), 969–996.
- Chetty, R., A. Looney, and K. Kroft (2009). Salience and taxation: Theory and evidence. *The American Economic Review* 99(4), 1145.
- Degan, A. and A. Merlo (2011). A structural model of turnout and voting in multiple elections. *Journal of the European Economic Association* 9(2), 209–245.
- Diermeier, D., H. Eraslan, and A. Merlo (2003). A structural model of government formation. *Econometrica* 71(1), 27–70.
- Diermeier, D. and A. Merlo (2004). An empirical investigation of coalitional bargaining procedures. *Journal of Public Economics* 88(3), 783–797.
- Diermeier, D., C. Prato, and R. Vlaicu (2015). Procedural choice in majoritarian organizations. *American Journal of Political Science* 59(4), 866–879.
- El Pais (2015). Clima social tras las elecciones del 24-M. [http://elpais.com/elpais/2015/06/19/media/1434733713\\_25358](http://elpais.com/elpais/2015/06/19/media/1434733713_25358).

- Europa Press (2007). Rajoy emula a Zapatero y se compromete a no gobernar si su partido no es el mas votado. <http://www.europapress.es/nacional/noticia-rajoy-emula-zapatero-compromete-no-gobernar-si-partido-no-mas-votado-20070904000108.html>.
- Europa Press (2011). El PP tendra la alcaldia de Olivenza, pueblo natal de Vara, dado que IU no apoyara a ningun otro grupo. <http://www.europapress.es/nacional/noticia-22-pp-tendra-alcaldia-olivenza-pueblo-natal-vara-dado-iu-no-apoyara-ningun-otro-grupo-20110609230840.html>.
- Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice* 50(1), 5–25.
- Finkelstein, A. (2009). E-ztax: Tax salience and tax rates. *The Quarterly Journal of Economics* 124(3), 969–1010.
- Flegenheimer, M. (2016). Ted Cruz, facing hard G.O.P. path, seeks a good argument. *NYTimes*, [https://www.nytimes.com/2016/04/22/us/politics/ted-cruz-campaign.html?\\_r=0](https://www.nytimes.com/2016/04/22/us/politics/ted-cruz-campaign.html?_r=0).
- Folke, O., T. Persson, and J. Rickne (2015). The primary effect: Preference votes and political promotions. *American Political Science Review*, forthcoming.
- Frechette, G., J. H. Kagel, and M. Morelli (2005). Nominal bargaining power, selection protocol, and discounting in legislative bargaining. *Journal of Public Economics* 89(8), 1497–1517.
- Hartzmark, S. M. (2015). The worst, the best, ignoring all the rest: The rank effect and trading behavior. *Review of Financial Studies* 28(4), 1024–1059.
- Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies* 79(3), 933–959.
- Krehbiel, K. (2004). Legislative organization. *The Journal of Economic Perspectives* 18(1), 113–128.
- Kuziemko, I., R. W. Buell, T. Reich, and M. I. Norton (2014). Last-place aversion: Evidence and redistributive implications. *The Quarterly Journal of Economics* 129(1), 105–149.
- Lacetera, N., D. G. Pope, and J. R. Sydnor (2012). Heuristic thinking and limited attention in the car market. *The American Economic Review* 102(5), 2206–2236.
- Laver, M. (1998). Models of government formation. *Annual Review of Political Science* 1(1), 1–25.
- Laver, M. and K. Benoit (2015). The basic arithmetic of legislative decisions. *American Journal of Political Science* 59(2), 275–291.
- Laver, M. and K. A. Shepsle (1990). Coalitions and cabinet government. *American Political Science Review* 84(3), 873–890.
- Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. *Journal of Econometrics* 142(2), 675–697.
- Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2), 281–355.
- Lizzeri, A. and N. Persico (2001). The provision of public goods under alternative electoral incentives. *American Economic Review* 91(1), 225–239.
- Magre-Ferran, J. and X. Bertrana-Horta (2005). Municipal presidentialism and democratic consolidation in Spain. *R. Berg, N. Rao (red.), Transforming Local Political Leadership, New York: Palgrave Macmillan*, 73–84.
- Merlo, A. (1997). Bargaining over governments in a stochastic environment. *Journal of Political Economy* 105(1), 101–131.
- Merlo, A. and C. Wilson (1995). A stochastic model of sequential bargaining with complete information. *Econometrica: Journal of the Econometric Society*, 371–399.

- Morelli, M. (1999). Demand competition and policy compromise in legislative bargaining. *American Political Science Review* 93(4), 809–820.
- Murray, M. (2016). Poll: 62% say republican with most votes should be nominee. *NBC News*, <http://www.nbcnews.com/politics/first-read/poll-62-say-republican-most-votes-should-be-nominee-n556846>.
- Ortoleva, P. and E. Snowberg (2015). Overconfidence in political behavior. *The American Economic Review* 105(2), 504–535.
- Osborne, M. J. and A. Rubinstein (1994). *A course in game theory*. MIT press.
- Passarelli, F. and G. Tabellini (2017). Emotions and political unrest. *Journal of Political Economy* 125(3), 903–946.
- Perraudin, F. (2015). Lib Dems would give biggest party time to form coalition, says Clegg. <https://www.theguardian.com/politics/2015/may/05/nick-clegg-let-largest-party-after-election-form-government>.
- Persson, T., G. Roland, and G. Tabellini (1997). Separation of powers and political accountability. *The Quarterly Journal of Economics* 112(4), 1163–1202.
- Persson, T. and G. E. Tabellini (2005). *The economic effects of constitutions*. MIT press.
- Pope, D. G. (2009). Reacting to rankings: The effect of electoral systems on voter turnout: Evidence from America’s best hospitals. *Journal of Health Economics* 28(6), 1154–1165.
- Ray, D. and R. Vohra (2014). Coalition formation. *Handbook of Game Theory* 4, 239–326.
- Sanz, C. (2017a). The effect of electoral systems on voter turnout: Evidence from a natural experiment. *Political Science Research and Methods* 5(4), 689–710.
- Sanz, C. (2017b). Increasing women’s representation in politics: The importance of the electoral system. *Unpublished Manuscript*.
- Silver, N. (2016). Why Republican voters decided on Trump. <https://fivethirtyeight.com/features/why-republican-voters-decided-on-trump/>.
- Snyder Jr, J. M., M. M. Ting, and S. Ansolabehere (2005). Legislative bargaining under weighted voting. *American Economic Review* 95(4), 981–1004.
- Sweeting, D. (2009). The institutions of strong local political leadership in Spain. *Environment and Planning C: Government and Policy* 27(4), 698–712.
- Vallés, J. M. and Q. Brugué (2001). El gobierno local. *Política y Gobierno en España*, 103–121.
- Warwick, P. V. and J. N. Druckman (2001). Portfolio salience and the proportionality of payoffs in coalition governments. *British Journal of Political Science* 31(04), 627–649.

Figure 1a: Effect of Being First (Instead of Second) Most Voted

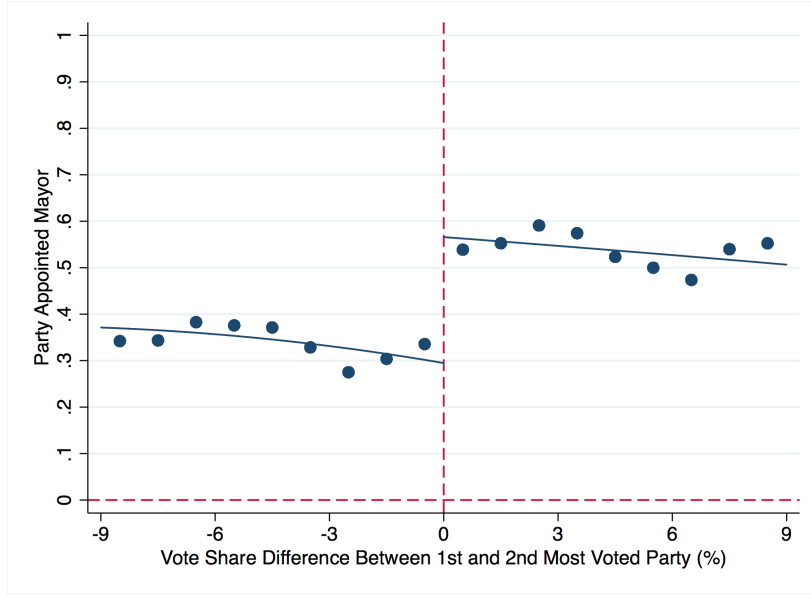
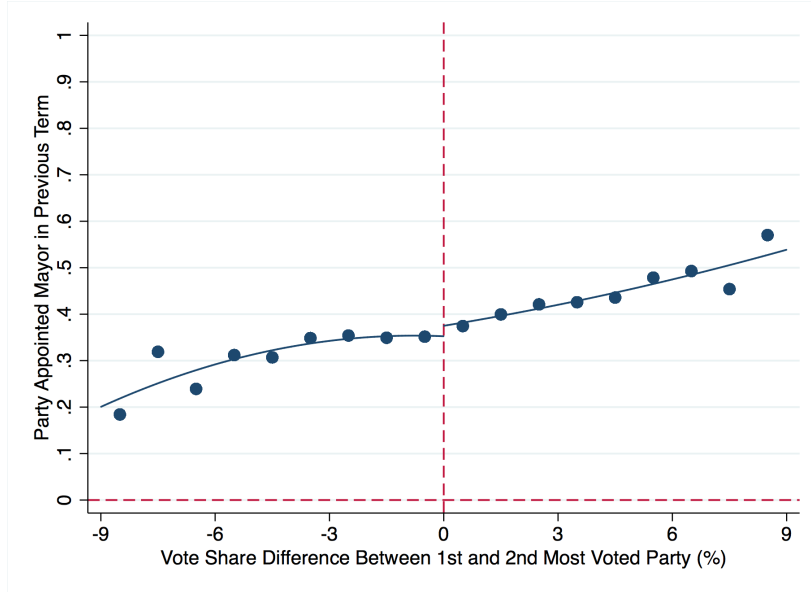


Figure 1b: Placebo Test: “Effect” of Most Voted on Lagged Outcome



The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties in elections in which they tied in seats. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative for the second most voted. Circles represent the local averages of a dummy indicating whether the party appointed the mayor (Panel A) or if the party appointed the mayor in the previous ( $t - 1$ ) term (Panel B). Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Figure 2a: Effect of Being Second (Instead of Third) Most Voted

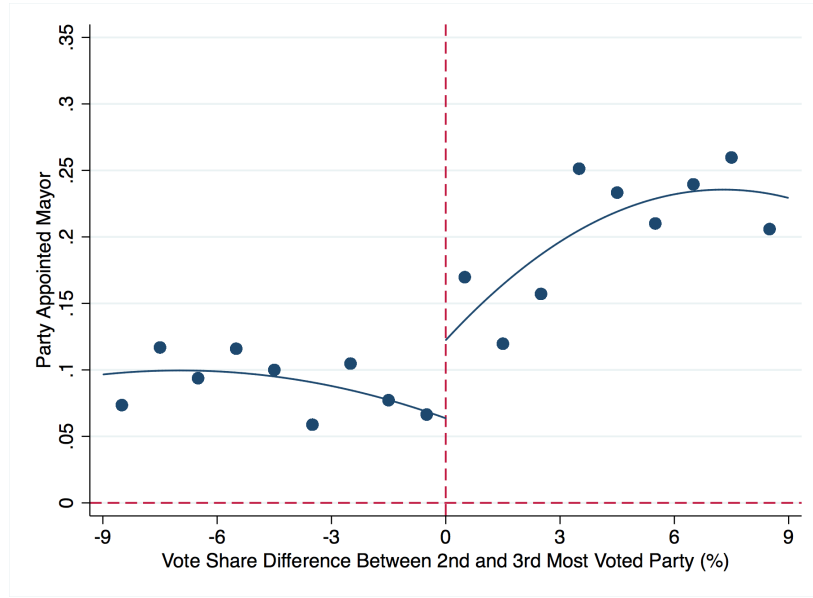
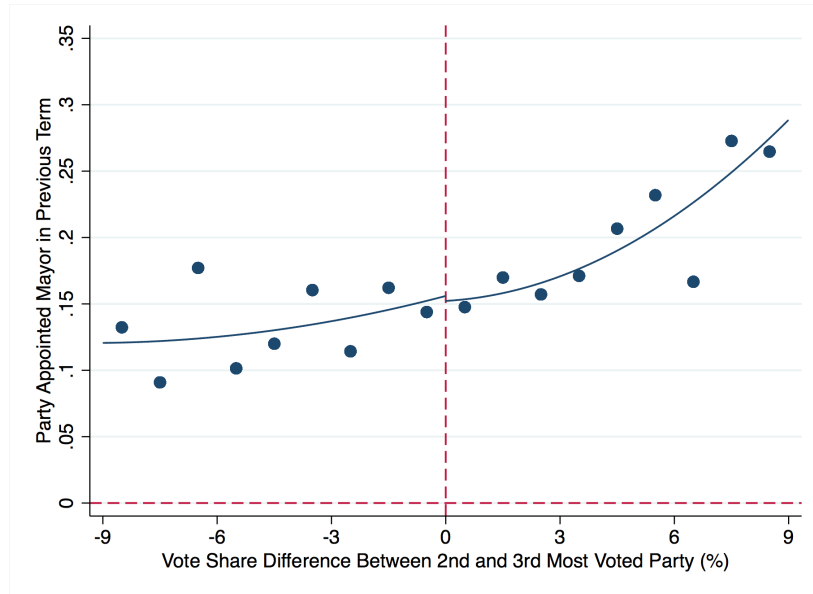
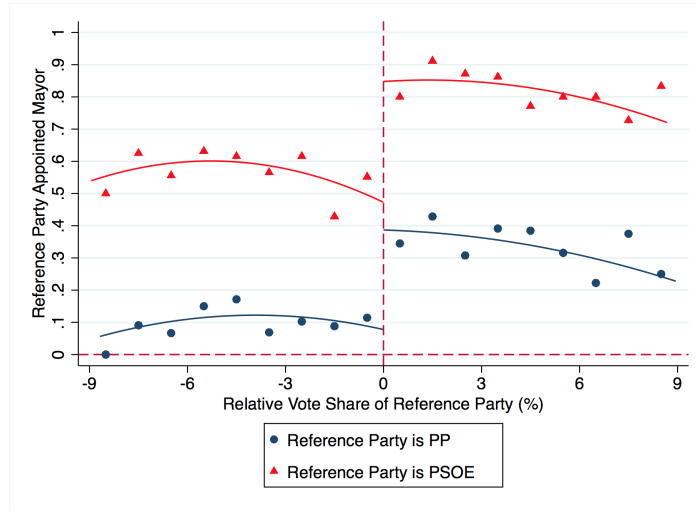


Figure 2b: Placebo Test: “Effect” of Second Most Voted on Lagged Outcome



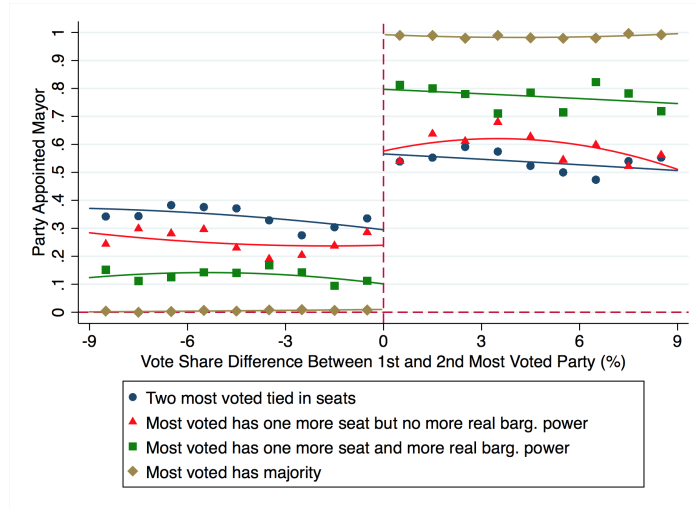
The unit of observation is a party-municipality-year. Sample is restricted to the second and third most voted parties in elections in which they tied in seats and the most voted party did not obtain a majority of seats. The running variable (horizontal axis) is the difference in vote shares between the second and third most voted parties: positive for the second most voted party and negative for the third most voted. Circles represent the local averages of a dummy indicating whether the party appoints the mayor (Panel A) or if the party appointed the mayor in the previous ( $t - 1$ ) term (Panel B). Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Figure 3: Effect of Being Most Voted: Cases with Left-Wing Majority



The unit of observation is a party-municipality-year. Sample is restricted to elections in which the *Partido Socialista Obrero Español (PSOE)* and the *Partido Popular (PP)* are the two most voted parties and the third most voted party is the *Izquierda Unida (IU)*. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties, taking either the PSOE or the PP as the reference party. Hence red triangles (blue circles) to the left of the vertical line at zero are cases where the PSOE (PP) was the second most voted party and, to the right, the most voted. Markers represent the local averages of a dummy indicating whether the party appoints the mayor. Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Figure 4: Effect of First Place by Legislature Type



The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties. Each plot restricts the sample to a different case of seat composition in the legislature. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative for the second most voted. Circles represent the local averages of a dummy indicating whether the party appoints the mayor. Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Figure 5a: Event Study for Effect of Mayoral Appointment on Vote Shares, by Party Rank

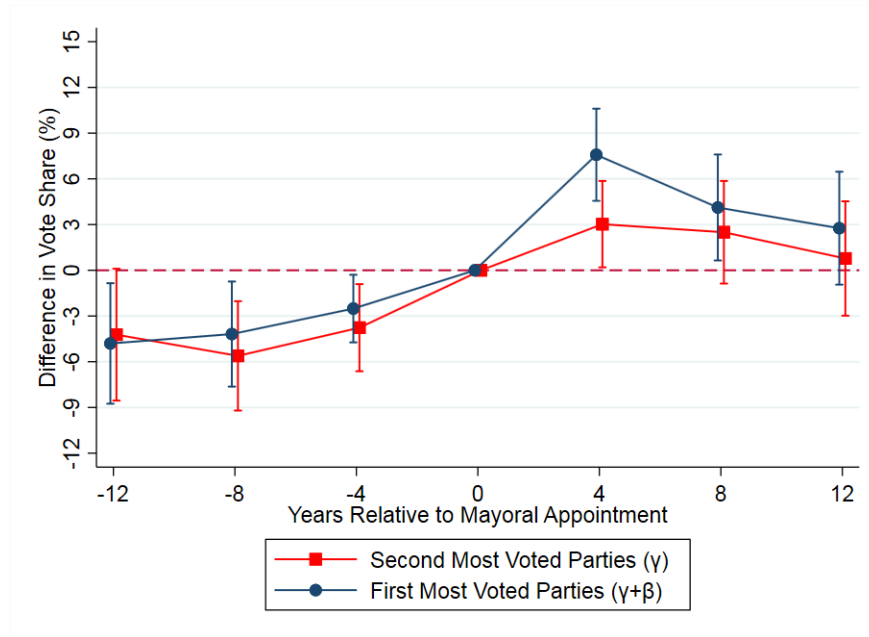
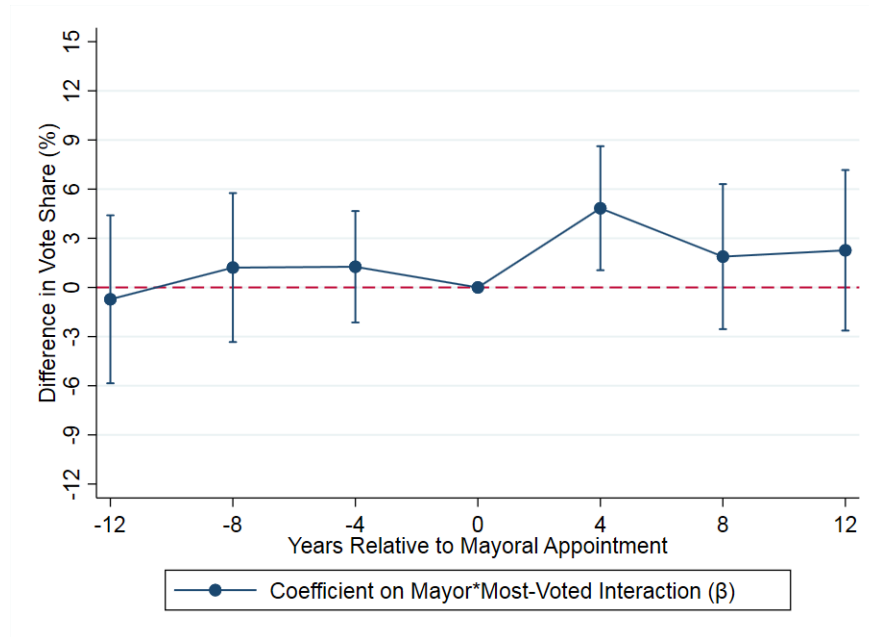


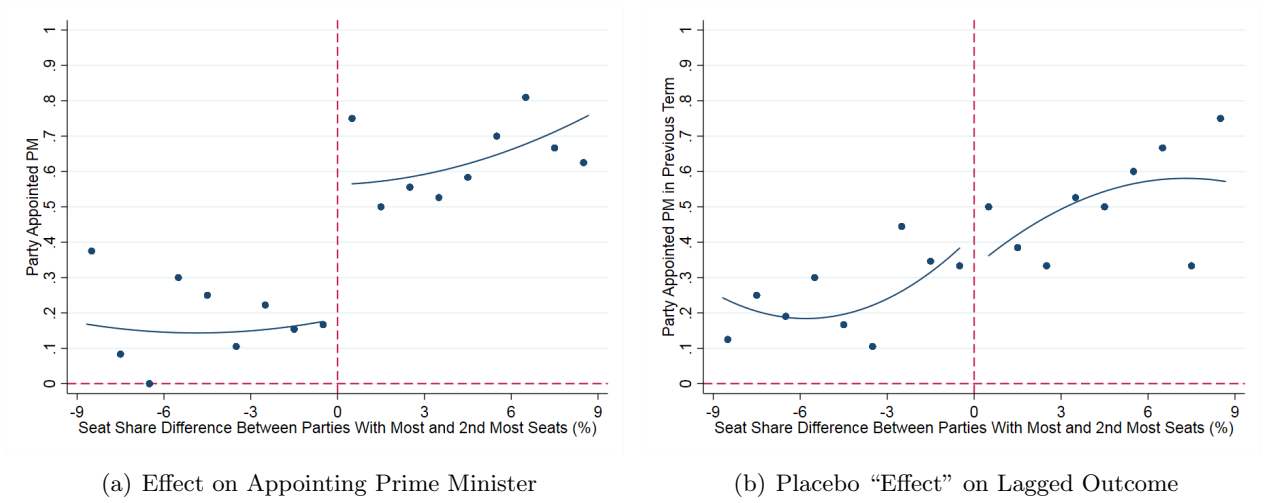
Figure 5b: Event Study for Effect of Mayoral Appointment on Vote Shares, Triple-Differences



Whiskers represent 95% confidence intervals based on standard errors clustered at the municipality level. Sample is restricted to elections in which the two most voted parties tied in seats and their difference in vote shares was less than 1% of the total vote. Vote shares are normalized to zero at  $t = 0$ . Red squares (blue circles) in Panel A show how the share of votes for a second-placed (first-placed) party that appoints a mayor at  $t = 0$  evolves relative to a second-placed (first-placed) party that does not, obtained by estimating  $\gamma$  and  $\gamma + \beta$  from equation (2) with different time horizons (see text for further details). Blue circles in Panel B represent the triple-difference event study: the difference between Panel A markers, obtained by estimating  $\beta$  from equation (2) with different time horizons (see text for further details).



Figure 6: European National Parliaments - Effect of Most Seats and Placebo Test



The unit of observation is a party-country-year. Sample is restricted to the two parties with the most seats in the parliament. The running variable (horizontal axis) is the difference in *seat* shares between the two parties with the most seats: positive with the most seats and negative for the party with the second most number of seats. Circles represent the local averages of a dummy indicating whether the party appoints the prime minister (Panel A) or if the party appointed the prime minister in the previous ( $t - 1$ ) term (Panel B). Averages are calculated within 1 p.p.-wide bins of seat share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Table 1: Effect of Being First (Instead of Second) Most Voted

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: Main Outcome</i>					
Party Appointed	0.353	0.185***	0.203***	0.295***	0.241***
Mayor		(0.059)	(0.044)	(0.037)	(0.046)
<i>N</i>		2028	876	5796	5796
<i>Panel B: Lagged Outcome (Placebo Test)</i>					
Party Appointed	0.358	0.011	0.023	-0.015	0.014
Mayor, $t - 1$		(0.046)	(0.040)	(0.034)	(0.043)
<i>N</i>		2714	876	5796	5796
<i>Panel C: Covariate Balance</i>					
Party is PP	0.310	0.006	0.009	-0.027	-0.036
		(0.041)	(0.039)	(0.034)	(0.042)
<i>N</i>		3088	876	5796	5796
Party is PSOE	0.405	0.0003	0.005	0.018	0.001
		(0.045)	(0.044)	(0.036)	(0.046)
<i>N</i>		3222	876	5796	5796
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to the two most voted parties in elections in which they tied in seats. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for a 2nd most voted party that tied with the most voted party (using the specification in column 1). Optimal bandwidths are based on [Imbens and Kalyanaraman \(2012\)](#), being equal to 2.32%, 3.19%, 3.75%, and 3.92% for the four dependent variables, respectively.

Table 2: Effect of Being Second (Instead of Third) Most Voted

Dependent Variable	3rd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: Main Outcome</i>					
Party Appointed	0.067	0.092**	0.103***	0.059**	0.073**
Mayor		(0.043)	(0.028)	(0.028)	(0.036)
<i>N</i>		888	542	3132	3132
<i>Panel B: Lagged Outcome (Placebo Test)</i>					
Party Appointed	0.149	0.004	0.004	-0.024	0.006
Mayor, $t - 1$		(0.037)	(0.034)	(0.034)	(0.045)
<i>N</i>		1844	542	3132	3132
<i>Panel C: Covariate Balance</i>					
Party is PP	0.312	-0.092	-0.044	-0.025	-0.033
		(0.072)	(0.043)	(0.043)	(0.056)
<i>N</i>		856	542	3132	3132
Party is PSOE	0.285	-0.028	-0.040	0.0043	-0.031
		(0.057)	(0.043)	(0.043)	(0.056)
<i>N</i>		1234	542	3132	3132
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to the second and third most voted parties in elections in which they tied in seats and the most voted party did not obtain a majority of seats. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 3rd-Place Mean is the estimated value of the dependent variable for a 3rd most voted party that tied with the 2nd most voted party (using the specification in column 1). Optimal bandwidths are based on [Imbens and Kalyanaraman \(2012\)](#), 1.69%, 3.96%, 1.63%, and 2.41% for the four dependent variables, respectively.

Table 3: Do Voters Punish Parties That Break the Norm? Triple-Difference Estimates

	Outcome ( $v_{pm,t+1} - v_{pm,t}$ )					Lagged Outcome ( $v_{pm,t} - v_{pm,t-1}$ )
	(1)	(2)	(3)	(4)	(5)	(6)
Mayor <sub>t</sub> *Most Voted <sub>t</sub> ( $\beta$ )	4.834** (1.931)	4.232** (1.881)	3.382* (1.814)	5.116* (2.909)	3.504 (2.512)	-1.263 (1.735)
Mayor <sub>t</sub> ( $\gamma$ )	2.868** (1.449)	3.091** (1.469)	2.937** (1.404)	1.175 (2.059)	4.998** (2.046)	3.822*** (1.441)
Most Voted <sub>t</sub> ( $\delta$ )	-1.693 (1.393)	-1.532 (1.355)	-1.399 (1.304)	-0.691 (2.035)	-2.017 (1.898)	-1.184 (1.121)
Constant	-1.160 (0.785)	-0.681 (3.025)	6.630 (5.646)	0.753 (1.055)	-3.372*** (1.163)	-0.877 (0.783)
Province effects		Y	Y			
Party effects			Y			
Only elections w. vote share of 3rd > median				Y		
Only elections w. vote share of 3rd < median					Y	
<i>N</i>	664	664	664	332	332	694

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-election. See discussion of equation (2) in text for specification. Outcome in columns (1)-(5) is the growth in vote share between the election immediately preceding a possible mayoral appointment ( $t$ ) and the next election ( $t + 1$ ). Outcome in column (6) is growth between time  $t$  and  $t - 1$  (a placebo test). The sample is restricted to elections in which the two most voted parties tied in seats and their difference in vote shares was less than 1% of the total vote.

Table 4: Effect of Being Most Voted on Appointing the Mayor, by Strength of Third-Placed Party

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: Third most voted party vote share above median</i>					
Party Appointed	0.290	0.306***	0.290***	0.364***	0.334***
Mayor		(0.078)	(0.056)	(0.048)	(0.062)
<i>N</i>		1014	468	2756	2756
<i>Panel B: Third most voted party vote share below median</i>					
Party Appointed	0.430	0.0402	0.103	0.214***	0.118*
Mayor		(0.088)	(0.066)	(0.059)	(0.070)
<i>N</i>		1014	408	3040	3040
<i>p</i> -value: test of equal effects		0.0228	0.0321	0.0498	0.0212
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to the two most voted parties in elections in which they tied in seats. In Panel A (Panel B), sample is further restricted to elections where the third-placed party has vote share above (below) the median of the sample used in column (1): 16.5%. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for a 2nd most voted party that tied with the 1st most voted party (using the specification in column 1). The optimal bandwidth is calculated based on the entire sample and is 2.32% (Imbens and Kalyanaraman 2012).

Table 5: National Parliaments Data: Effect of Having Most Seats on Appointing Prime Minister

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: Main Outcome</i>					
Party Appointed	0.202	0.303**	0.583**	0.387***	0.387**
Prime Minister		(0.130)	(0.210)	(0.122)	(0.145)
<i>N</i>		224	24	504	504
<i>Panel B: Lagged Outcome (Placebo Test)</i>					
Party Appointed	0.414	-0.0316	0.167	0.0852	0.136
Prime Minister, $t - 1$		(0.163)	(0.332)	(0.122)	(0.129)
<i>N</i>		152	24	504	504
<i>Panel C: Covariate Balance (Outcome Predicted from Party Ideology)</i>					
Party Appointed	0.417	-0.021	-0.035	-0.015	-0.009
PM (Predicted)		(0.028)	(0.078)	(0.026)	(0.033)
<i>N</i>		222	24	504	504
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the country level in parentheses. The unit of observation is a party-country-year. The sample is restricted to the two parties with the most seats in the parliament. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for the party with the 2nd-most seats that tied with the party with most seats (using the specification in column 1). Optimal bandwidths are based on [Imbens and Kalyanaraman \(2012\)](#), being equal to 7.39%, 4.82%, and 7.36%, for the three dependent variables, respectively. See text for the construction of the outcome on Panel C.



## Online Appendices - Not for Publication

### Appendix A: Example of Voter Belief Updating in the Theoretical Framework

To illustrate the workings of the model, this section provides an example using a specific distribution of possible states of the world. Recall that  $\mathbf{p}_t = [p_t^A, p_t^B, p_t^C]$  denotes the probabilities of state  $s_t \in \{A, B, C\}$  occurring. In particular, assume that  $G(\mathbf{p}_t)$  is such that its possible realizations are i)  $\mathbf{p}_t = [0.45, 0.35, 0.20]$ , ii)  $\mathbf{p}_t = [0.35, 0.45, 0.20]$ , and iii)  $\mathbf{p}_t = [0.20, 0.20, 0.60]$ . Each realization can occur with probability equal to  $1/3$ .

Hence, when a period starts, voters have priors that each of the states of the world are equally likely. After observing a signal  $\sigma$  equal to  $A$ , a voter updates and then believes that the probability that realization (i) occurred is  $\frac{0.45 \cdot (1/3) \cdot (0.45 + 0.35 + 0.2)}{1/3} = 0.45$ . She similarly believes that the probability that (ii) occurred is 0.35 and that (iii) occurred is 0.2.

Hence, observing signal  $A$  makes her update that the probability of each state occurring:  $\Pr(s_t = A | \sigma_t = A) = 0.45^2 + 0.35^2 + 0.2^2 = 0.365$ ;  $\Pr(s_t = B | \sigma_t = A) = 0.45 \cdot 0.35 + 0.35 \cdot 0.45 + 0.2^2 = 0.355$ ; and  $\Pr(s_t = C | \sigma_t = A) = 0.45 \cdot 0.2 + 0.35 \cdot 0.2 + 0.2 \cdot 0.6 = 0.280$ . Similarly, observing a signal  $B$  will make her believe that state  $B$  has a 0.365 probability of occurring (while probability of  $A$  and  $C$  are 0.355 and 0.280, respectively). A similar calculation yields the updated beliefs after a voter observes signal  $C$ :  $\Pr(s_t = A | \sigma_t = C) = \Pr(s_t = B | \sigma_t = C) = 0.280$  and  $\Pr(s_t = C | \sigma_t = C) = 0.440$ .

Note that this distribution satisfies the  $\Pr(s_t = i | \sigma_t = i) > \Pr(s_t = j | \sigma_t = i)$  for all  $i \neq j$  condition. So a voter that observes signal  $i$  prefers party of type  $i$  to be the mayor. However, if all voters vote according to their signals, the actual vote shares will match one of the  $\mathbf{p}$  realizations - e.g., if realization (i) occurs, the vote shares of parties  $A$ ,  $B$ , and  $C$  will be 0.45, 0.35, and 0.20, respectively. After observing such vote shares, citizens would then update accordingly: e.g., expect that the probability that the state is  $A$  is 0.45. This implies that then *all* voters will prefer party  $A$  to appoint the mayor, but at this point representation in the legislature is already determined. Given that party  $A$  does not have a majority of the votes, it might be possible for  $B$  or  $C$  to appoint the mayor. This illustrates the main conflict between voters and parties in the model. Vote shares aggregate diffuse information from the voters, which informs which party they prefer would appoint the mayor. However, after the election takes place, the decision on which party appoint the mayors may not necessarily heed to the preferences of voters.

Since  $G(\mathbf{p}_t)$  is assumed to be i.i.d. and serially uncorrelated, when a new period starts all voters beliefs about the state of the world return to the same prior (so past election results and mayoral appointments do not inform their rule). Lastly, in this particular example, a near tie between two parties for the most voted position is not possible. However, if realizations (i) and (ii) of the  $G(\mathbf{p}_t)$

were instead  $[0.40 + \epsilon, 0.40 - \epsilon, 0.20]$  and  $[0.40 - \epsilon, 0.40 + \epsilon, 0.20]$ , with  $\epsilon \rightarrow 0$ , we have a case where parties  $A$  and  $B$  almost tie and the updating works similarly.

## Appendix B: Alternative Bargaining Procedures

The theoretical framework assumes a specific bargaining procedure for mayoral appointments. Beyond tractability and simplicity, our particular choice of assumptions is also made to better match the one round of voting present in the Spanish context. This appendix discusses how the results in Section 4 are robust to a different bargaining procedure, which allows for both infinite rounds of bargaining and for rents from office to be divisible across parties.

This alternative procedure is inspired on [Baron and Ferejohn \(1989\)](#). As before, if one party obtains a majority, it can choose the allocation of rents. If no party has a majority, then one is randomly recognized to propose a division of the rents. Recognition probabilities are the same for all three parties. The non-recognized parties can accept or not this proposal. If at least one (non-recognized) party accepts, the recognized party appoints the mayor and the proposed division is realized.

If no party accepts, another identical round of bargaining begins, with another independent draw of the proposing party. Note that we do not need to specify a status quo appointment in this game, and technically the bargaining can last forever if offers are never accepted. Additionally, we abstract from discounting across bargaining rounds (so not to confuse with discounting across periods), however it is straightforward to incorporate them.

If no party has a majority, this bargaining game has an equilibrium with stationary (history independent) and symmetric strategies with the proposer offering  $x_t = 1/3$  to one randomly chosen party and  $x_t = 2/3$  for itself, with the first proposal being accepted.<sup>69</sup>

Proposition 1 can be adapted to, when no party obtains a majority, having all parties propose keeping  $2/3$  of the rents and offering  $1/3$  to another (randomly chosen) party. All parties accept such proposal. Note this implies that all parties have equal probability of appointing the mayor. If one party has a majority, then it appoints the mayor with certainty.

Proposition 2 can be similarly adapted. Note that we now equate “appointing the mayor” with “having a proposal accepted.” The new equilibrium strategy for a most voted party is: i) if recognized, offer to keep all the rents to itself; ii) if not recognized, to reject all offers. The equilibrium strategy for second and third most voted parties is: i) if recognized, offer to keep all rents to itself; ii) if not recognized, accept any offer that assigns it non-zero rents. If the offer assigns it zero rents, accept if it is from the most voted party and reject if it is from the second and third most voted.

These are clearly best responses to the most voted party. The strategies for second and third most voted parties are best responses given that they are indifferent between accepting or rejecting an offer that assigns zero rents. A deviation where they make an offer that is accepted (off the equilibrium path) cannot be a best response. Such deviation yields at most a payoff of one, since the party is never re-elected again, which is less than the continuation value of being re-elected given conditions (i)-(iii).

---

<sup>69</sup>A proposer keeps  $y$  and offers one randomly drawn other party  $1 - y$ . For the other party to accept, its payoff must be  $1 - y > V$ , where  $V$  is the continuation value of this legislative bargaining game. The proposer thus optimally makes this inequality bind, so the proposal is accepted. Hence, the continuation value equals  $V = \frac{1}{3}y + \frac{2}{3}(1 - y) = \frac{1}{3}$  and  $y = \frac{2}{3}$ .

This adapted version of Proposition 2 also leads to the most voted party appointing the mayor every period. Interestingly, the party appointing the mayor in the equilibrium described in Proposition 2 obtains more rents ( $x = 1$ ) than the one described in Proposition 1 ( $x = 2/3$ ).

## Appendix C: The Role of the Status Quo Rule in a Voting Game

To illustrate why the status quo rule is unlikely to play an important role in explaining the empirical results, we analyze a game matching the rules and incentives that parties face in our sample of Spanish municipalities. We focus on the case of a legislature with three parties, in which any two can form a majority. This case matches 90% of our sample, as discussed in Section 2.

Consider a game with three parties ( $A$ ,  $B$ , and  $C$ ), indexed by  $i$ . As in Spanish municipal councils, each party has only one candidate for mayor, which we also label  $A$ ,  $B$ , and  $C$ . Party preferences over the mayor are  $u_A(A) > u_A(B) > u_A(C)$ ,  $u_B(B) > u_B(C) > u_B(A)$ , and  $u_C(C) > u_C(B) > u_C(A)$ . This describes a situation in which two parties ( $B$  and  $C$ ) are ideologically aligned. Each party prefers to appoint the mayor itself. For parties  $B$  and  $C$ , their second option is the aligned party, and their least preferred option is  $C$ . While we assume party  $A$  prefers  $B$  over  $C$ , this is not crucial to the results. The strategy space is  $\{a, b, c, \phi\}$ . Parties can vote for any of the parties or abstain. Matching a situation where  $A$  and  $B$  are tied in seats and  $C$  has the same or fewer seats than  $A$  and  $B$ , if any two of the three parties vote for the same party  $i$ , then party  $i$  appoints the mayor.

For concreteness, we can think of  $A$  as the PP,  $B$  as the PSOE, and  $C$  as the IU. The leftist PSOE and IU prefer one of them to appoint the mayor over the right-wing PP. This case is depicted in Figure 3. Focusing when the PP is the reference party, assuming  $A$  is the most voted approximates the blue circles to the right of the cutoff, and the case where  $A$  is the second most voted the blue circles to the left of the cutoff. The question we address is whether the observed jump can be explained by  $A$  (the PP) changing to being the status quo as it crosses the cutoff.

Assume that  $A$  has a status quo status: it obtains the mayor if no party obtain two or more of the votes. In this case there are 14 Nash equilibria in pure strategies in the described game. Letting  $(s_A, s_B, s_C)$  denote equilibrium strategies, these are  $(a, a, a)$ ,  $(a, \phi, a)$ ,  $(a, a, \phi)$ ,  $(a, \phi, \phi)$ ,  $(a, b, b)$ ,  $(a, c, c)$ ,  $(b, b, b)$ ,  $(c, c, c)$ ,  $(\phi, a, a)$ ,  $(\phi, \phi, a)$ ,  $(\phi, a, \phi)$ ,  $(\phi, \phi, \phi)$ ,  $(\phi, b, b)$ , and  $(\phi, c, c)$ . Out of these 14 Nash equilibria,  $A$  appoints the mayor in eight,  $B$  in three, and  $C$  in three.

This multiplicity of equilibria is standard in voting games. However, the eight equilibria where  $A$  appoints the mayor are not robust to either i) allowing a coordinated deviation by two parties (e.g., strong Nash equilibrium or coalition-proofness) or ii) focusing on cases where parties do not play a weakly dominated strategy (e.g., trembling hand perfection or dominance solvability).

First, none of the eight equilibria where  $A$  appoints the mayor are robust to allowing both  $B$  and  $C$  to make a coordinated deviation.  $A$  appointing the mayor is the worst outcome for parties  $B$  and  $C$ . If the jointly deviate to either both voting for  $B$  or both for  $C$ , they can increase their utility. Note that while Nash equilibria only considers unilateral deviations, coordinated deviations seems like a natural case in a council with only three parties represented and where they can communicate before voting.

Second, note that actions  $b$  and  $c$  are weakly dominated (by  $A$  and  $\phi$ ) for party  $A$ .<sup>70</sup> Additionally,  $a$  and  $\phi$  are dominated by  $b$  (for party  $B$ ) and  $c$  (for party  $C$ ).<sup>71</sup> Hence, there are only four Nash equilibria where a party is not playing a weakly dominated strategy:  $(a, b, b)$ ,  $(a, c, c)$ ,  $(\phi, b, b)$ , and  $(\phi, c, c)$ . In none of them, party  $A$  appoints the mayor, even though it is the status quo. Assuming parties choose a weakly dominated action in a small voting game is unattractive. For example, trembling hand perfection eliminates all equilibrium with weakly dominated strategies: as long as party  $i$  believes that, even with a very small probability, one other party will vote for  $i$ , it will not be a best response to follow a weakly dominated strategy.

To illustrate the irrelevance of the status quo status further, one could reanalyze the game but now making party  $B$  the status quo. Following a similar argument, it can be shown there is no equilibrium where  $A$  appoints the mayor and players do not choose a weakly dominated strategy. Hence,  $A$ 's ability to appoint the mayor is unaffected by whether it is the status quo or not.

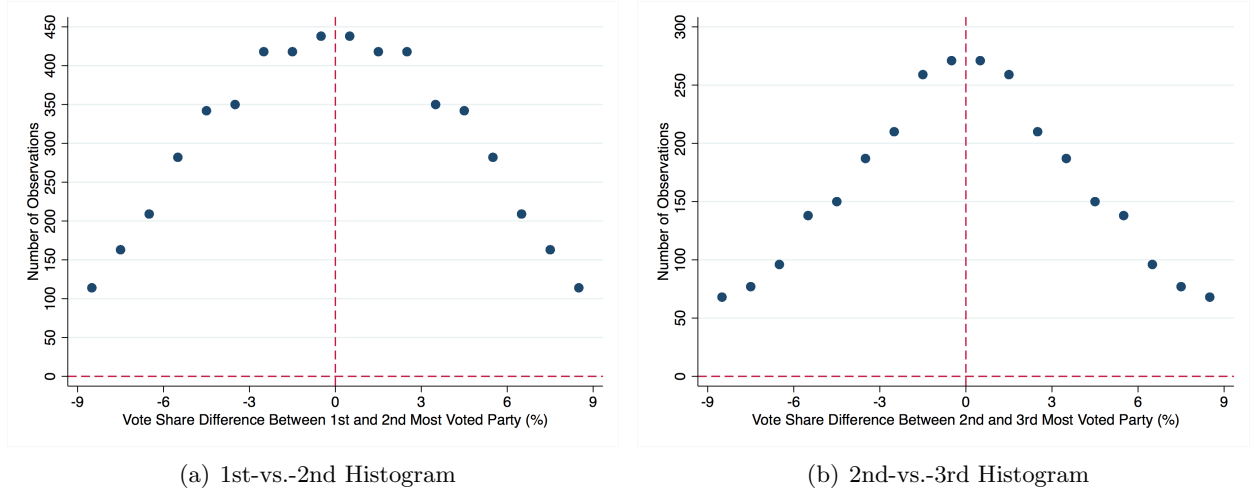
---

<sup>70</sup>If the other two parties are not casting the same vote,  $A$  and  $\phi$  can guarantee the best outcome for  $A$ . If only one other party is playing  $a$  or  $\phi$ , those actions are strictly better than  $b$  and  $c$  for  $A$ . If the other two parties are both voting  $b$  or  $c$ ,  $A$  is indifferent between all actions.

<sup>71</sup> $b$  guarantees the best outcome for  $B$  as long as one other party is playing  $b$ . If only one other party is playing  $b$ ,  $b$  is strictly better than any other action for  $B$ . If no other party is voting for  $b$ ,  $B$  is indifferent between  $b$ ,  $a$ , and  $\phi$ . A similar argument applies to party  $C$ .

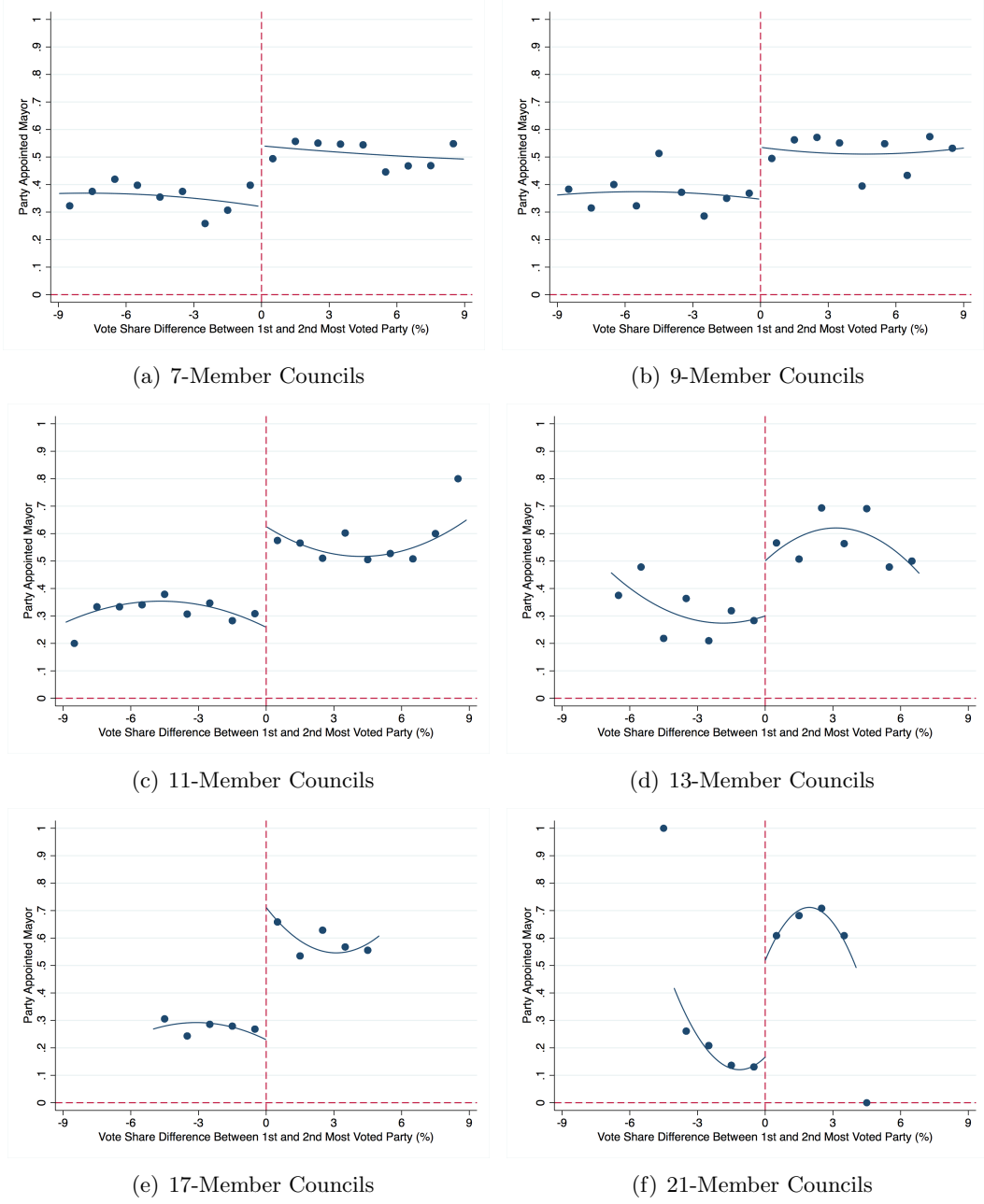
## Appendix Figures and Tables

Figure A1: Histograms: Distribution of Running Variable



The unit of observation is a party-municipality-year. Each circle in Panel A (Panel B) represents the number of observations in the respective circle on Figure 1 (Figure 2) of the main text. Panel A (Panel B) restricts the sample to the two most voted (second and third most voted) parties in elections in which they tied in seats. Panel B further restricts the sample to elections where the most voted party did not obtain a majority of seats. The running variable (horizontal axis) is the difference in vote shares between the first and second (Panel A) or second and third (Panel B) most voted parties. Circle represents the number of observations in each 1 p.p.-wide bin of vote share difference.

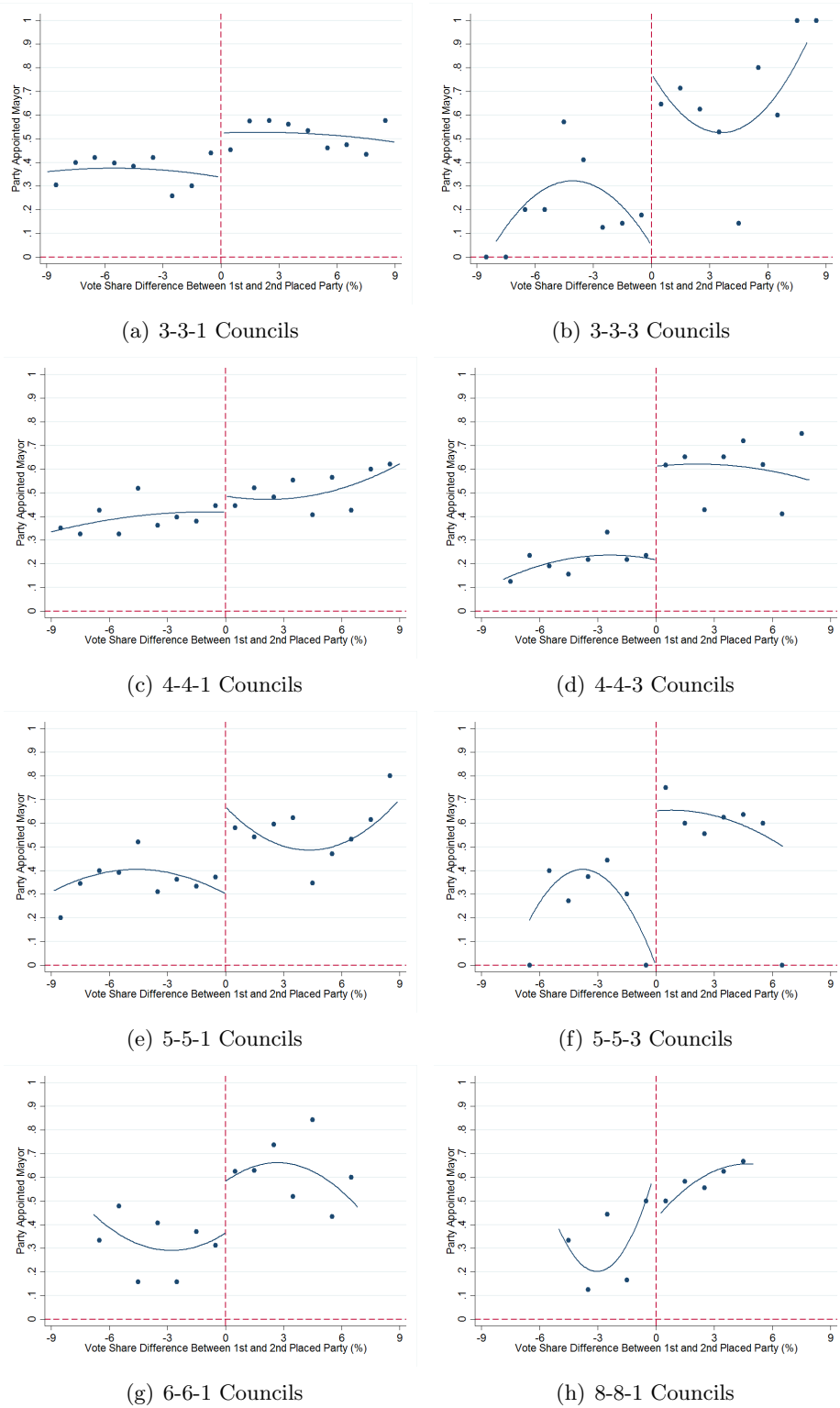
Figure A2: Effect of Being Most Voted, Heterogeneity by Council Size



The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties in elections in which they tied in seats. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative for the second most voted. Circles represent the local averages of a dummy indicating whether the party appointed the mayor. Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data. Each panel restricts the sample to elections with a specific council size. We report plots for all council sizes with a sample of at least 150 observations (75 elections).

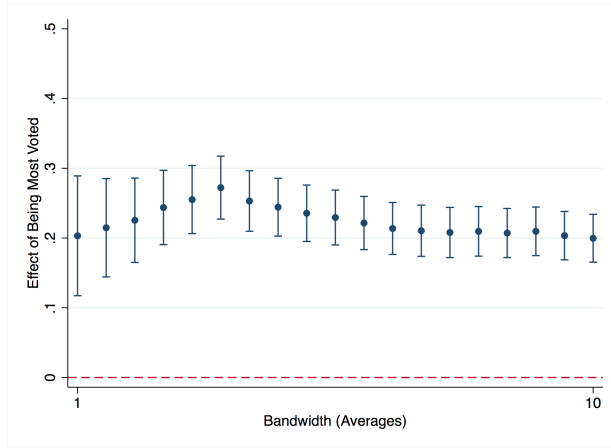


Figure A3: Effect of First-Place, by Council Type

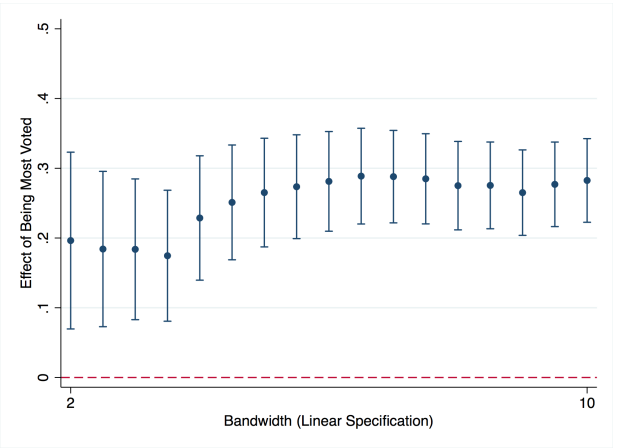


Notes are same to those on Figure A2, except each panel restricts the sample to councils with a particular seat configuration. We report plots for all configurations with a sample of at least 90 observations (45 elections).

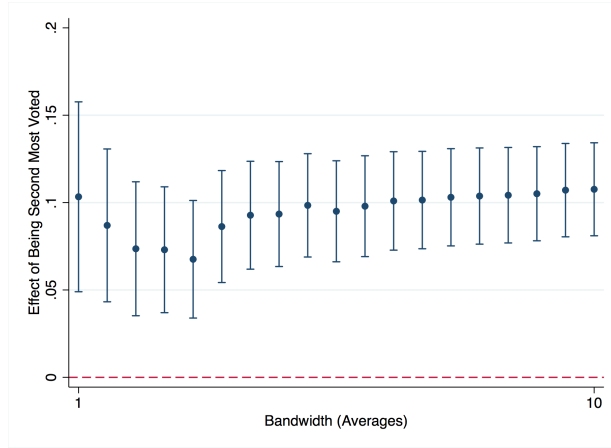
Figure A4: Robustness to Bandwidth Choice.



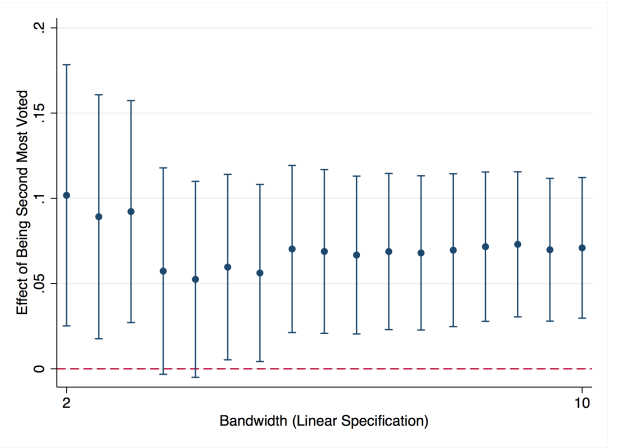
(a) Effect of 1st vs. 2nd (Specification: Means)



(b) Effect of 1st vs. 2nd (Specification: Linear)



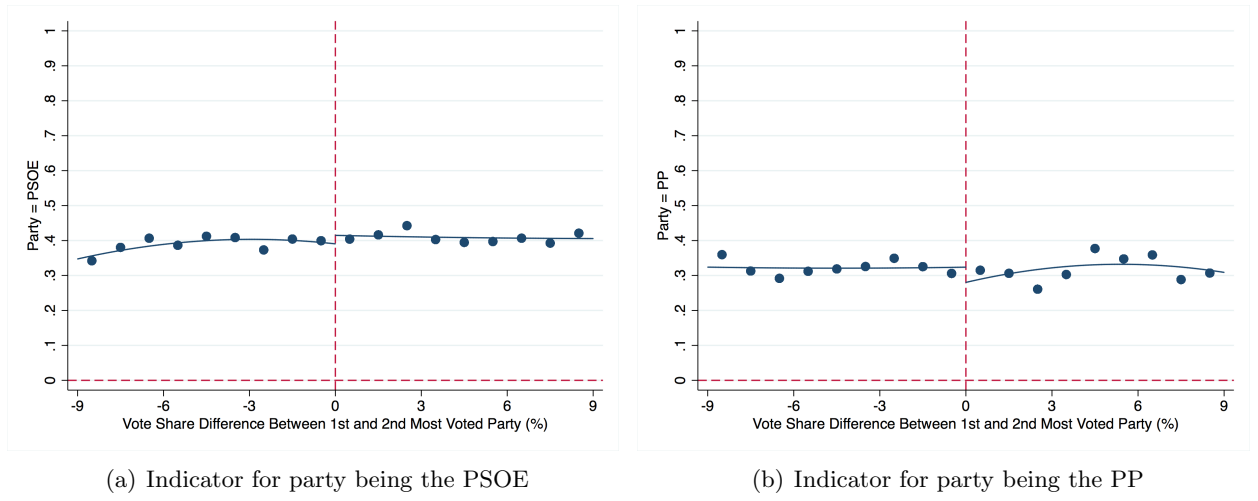
(c) Effect of 2nd vs. 3rd (Specification: Means)



(d) Effect of 2nd vs. 3rd (Specification: Linear)

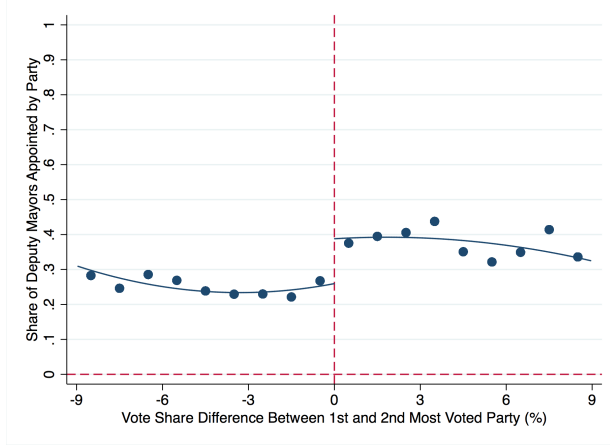
Circles represent estimated effects, using different bandwidth choices (horizontal axis). Whiskers represent the 95% confidence interval based on standard errors clustered at the municipality level.

Figure A5: Covariate Balance (1st vs. 2nd): Placebo “Effect” on Party Identity

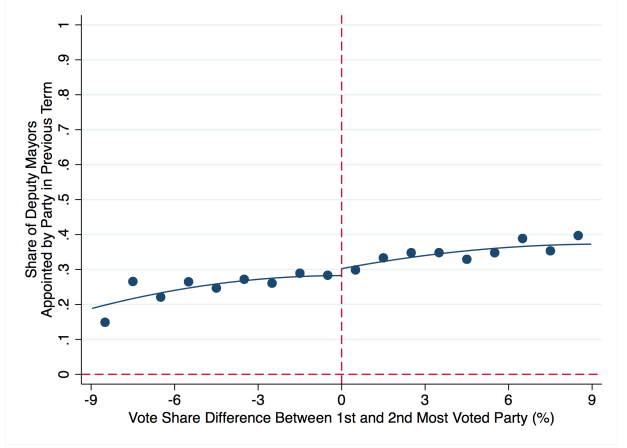


The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties in elections in which they tied in seats. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative for the second most voted. Circles represent the local averages of a dummy indicating whether the observation’s party is the *Partido Socialista Obrero Español* (Panel A) or *Partido Popular* (Panel B). Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

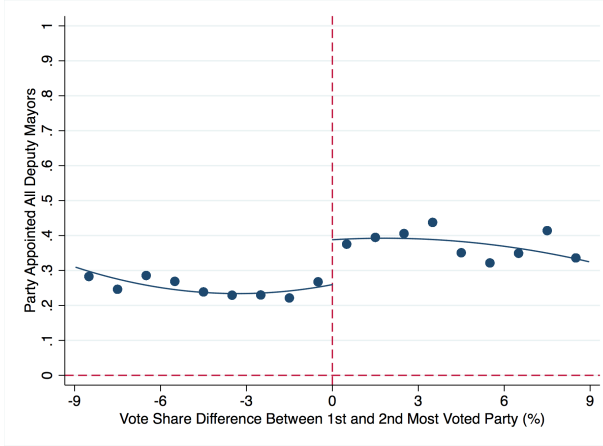
Figure A6: Effect of Being Most Voted on Deputy Mayors' Allocation



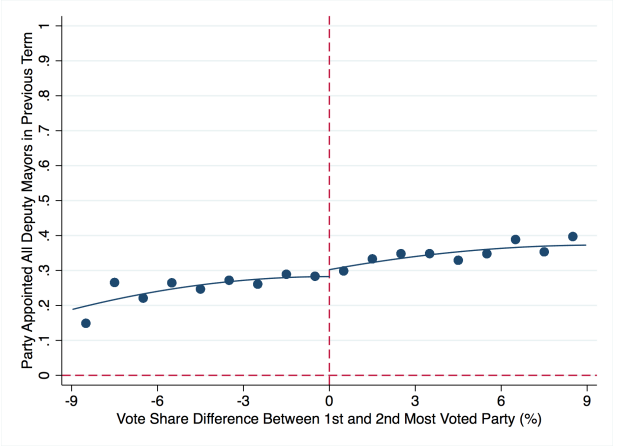
(a) Effect of Most Voted on Share of Deputy Mayors



(b) Placebo Test: “Effect” of Most Voted on Lagged Share of Deputy Mayors



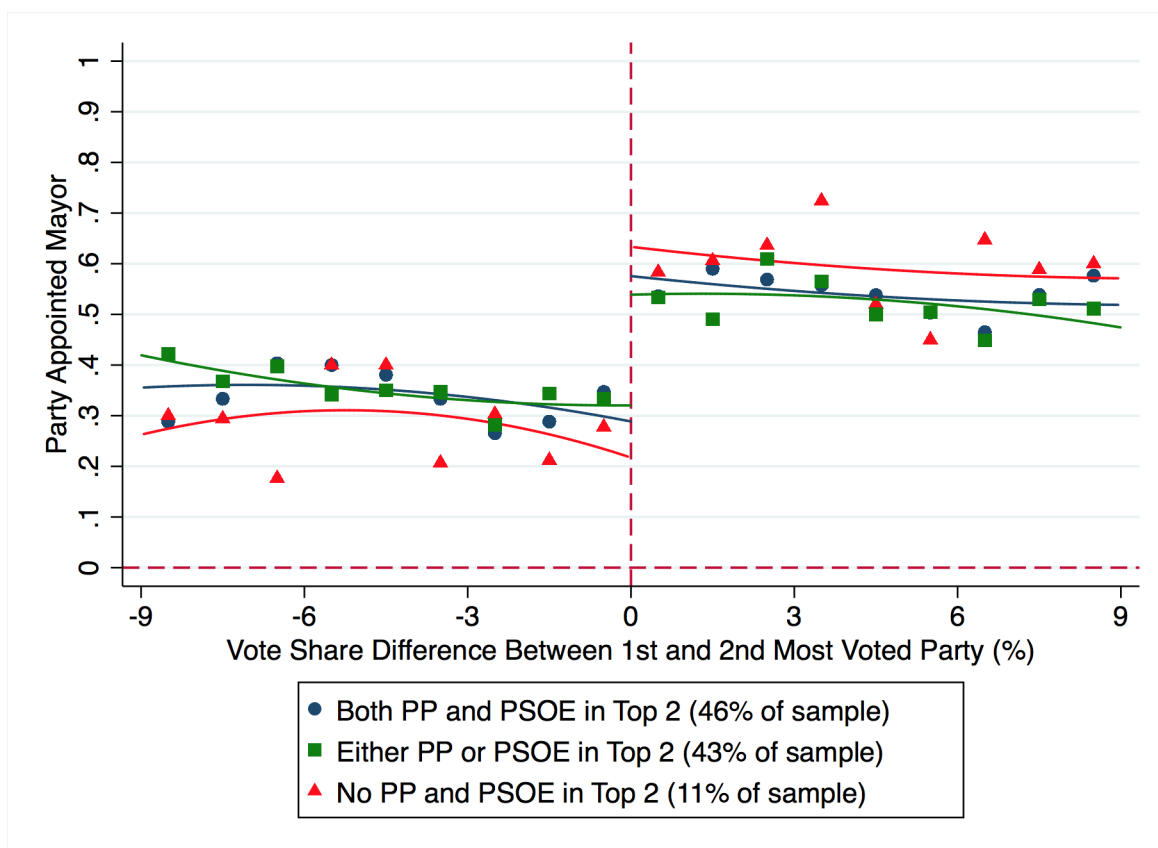
(c) Effect of Most Voted on Indicator for Appointing all Deputy Mayors



(d) Placebo Test: “Effect” of Most Voted on Indicator for Appointing all Deputy Mayors

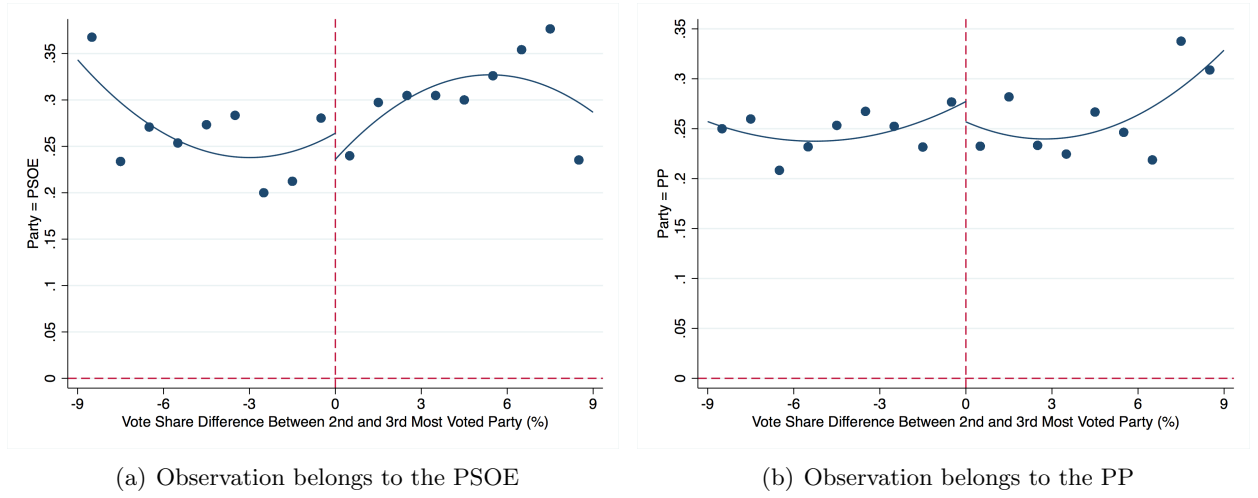
The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties in elections in which they tied in seats. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative for the second most voted. Circles in Panel A and B represent the local averages of the share of deputy mayors belonging to the party (Panel A) or that belonged to the party in the previous ( $t - 1$ ) term (Panel B). Circles in Panel C and D represent the local averages of an indicator for all deputy mayors belonging to the party (Panel C) or all having belonged to the party in the previous ( $t - 1$ ) term (Panel D). Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Figure A7: Effect of Being Most Voted: Heterogeneity by Party Identity



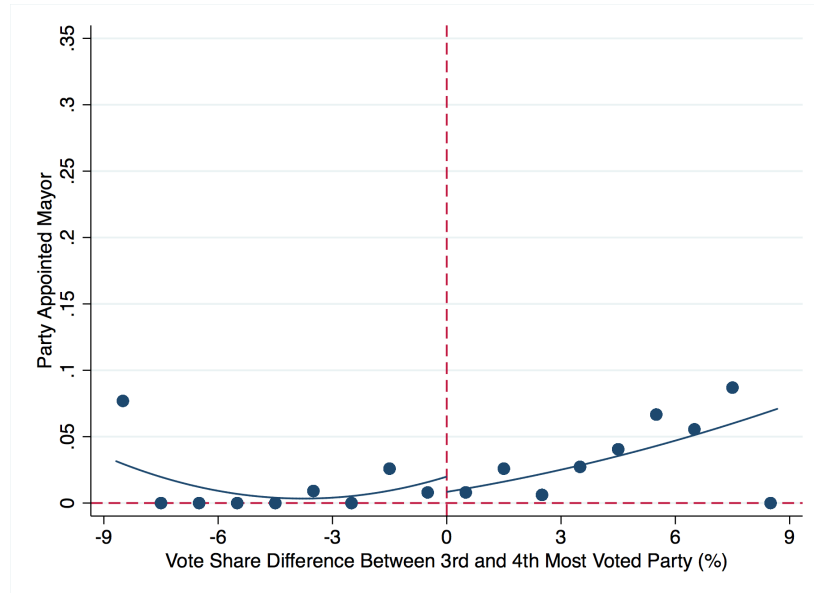
The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties in elections in which they tied in seats. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative for the second most voted. Markers represent the local averages of a dummy indicating whether the party appoints the mayor. Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data. The three separate plots are for the cases where both, either, or neither the *Partido Popular* (PP) and/or the *Partido Socialista Obrero Español* (PSOE) are amongst the two most voted parties.

Figure A8: Covariate Balance (2nd vs. 3rd): Placebo “Effect” on Party Identity



The unit of observation is a party-municipality-year. Sample is restricted to the second and third most voted parties in elections in which they tied in seats and the most voted party did not obtain a majority of seats. The running variable (horizontal axis) is the difference in vote shares between the second and third most voted parties: positive for the second most voted party and negative for the third most voted. Circles represent the local averages of a dummy indicating whether the observation’s party is the *Partido Socialista Obrero Español* (Panel A) or *Partido Popular* (Panel B). Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

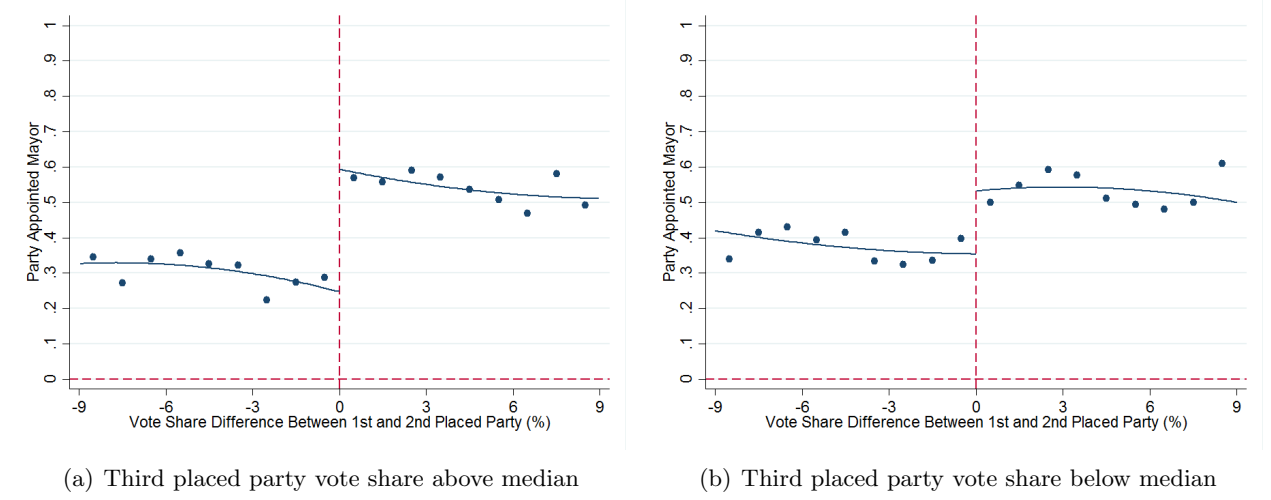
Figure A9: Effect of Being Third Most Voted: Third versus Fourth Place



The unit of observation is a party-municipality-year. Sample is restricted to the third and fourth most voted parties in elections in which they tied in seats and the most voted party did not obtain a majority of seats. The running variable (horizontal axis) is the difference in vote shares between the third and fourth most voted parties: positive for the third most voted party and negative for the fourth most voted. Circles represent the local averages of a dummy indicating whether the party appoints the mayor. Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

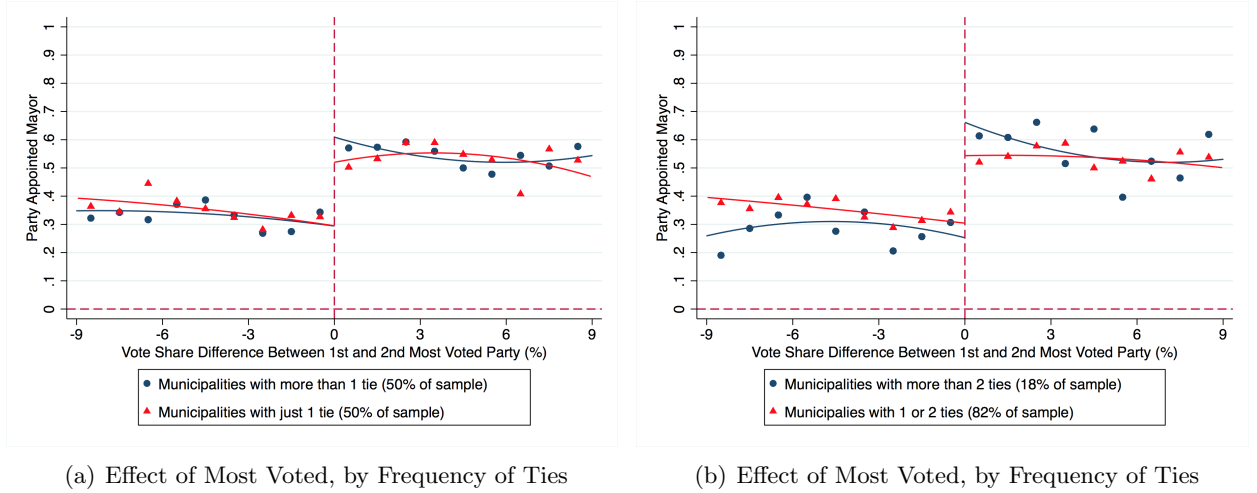


Figure A10: Effect Heterogeneity by Third-Placed Party Vote Share



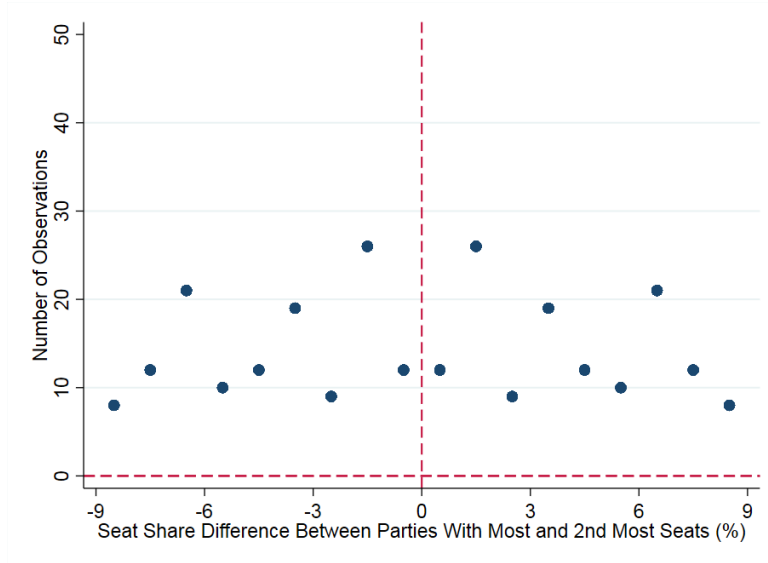
The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties in elections in which they tied in seats. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative for the second most voted. Circles represent the local averages of a dummy indicating whether the party appoints the mayor. Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data. In Panel A (Panel B), sample is further restricted to elections where the third most voted party has vote share above (below) the median of the sample used in column (1) in Table 1: 16.5%.

Figure A11: Effect Heterogeneity by Frequency of Ties in Municipality



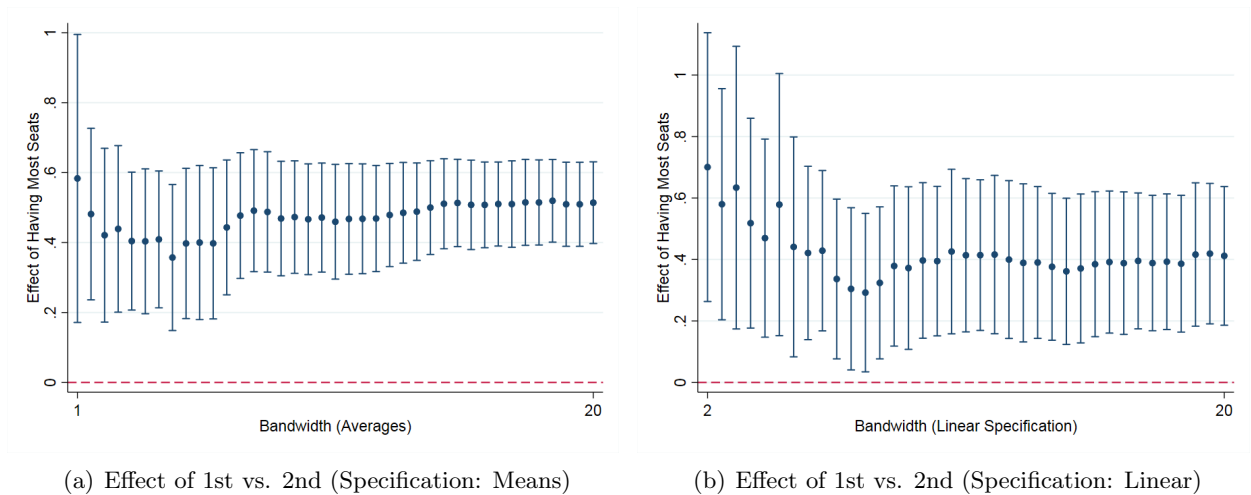
The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties in elections in which they tied in seats. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative for the second most voted. Markers represent the local averages of a dummy indicating whether the party appoints the mayor. Averages are calculated within 1 p.p.-wide bins of vote share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data. In Panel A, red triangles restrict the sample to municipalities with only one occurrence of the two most voted parties tying in seats in the sample period. Blue circles restrict the sample to municipalities where more than one tie in seats occurred. In Panel B, red triangles restrict the sample to municipalities where more than two ties occurred in the sample period, while the blue circles restrict it to cases where one or two ties occurred. See text for further details.

Figure A12: Histogram - National Parliaments Data



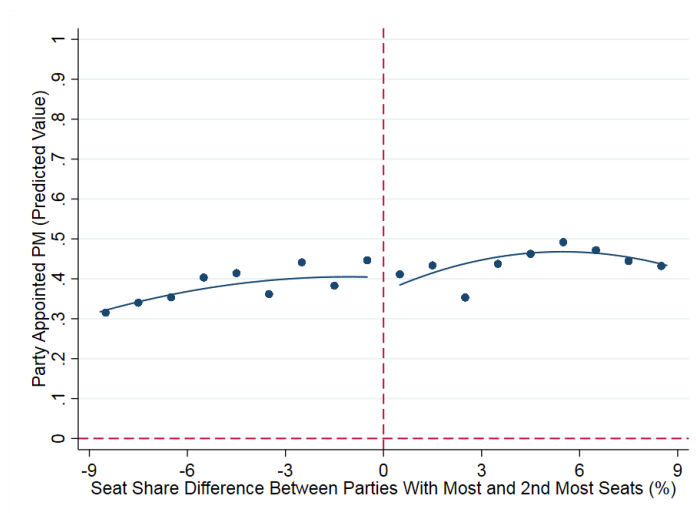
The unit of observation is a country-election-party. Each circle in represents the number of observations in the respective circle on Figure 6 of the main text. Sample is restricted to the the two most voted parties. Circles represents the number of observations in each 1 p.p.-wide bin of seat share difference.

Figure A13: Robustness to Bandwidth Choice - National Parliaments Data



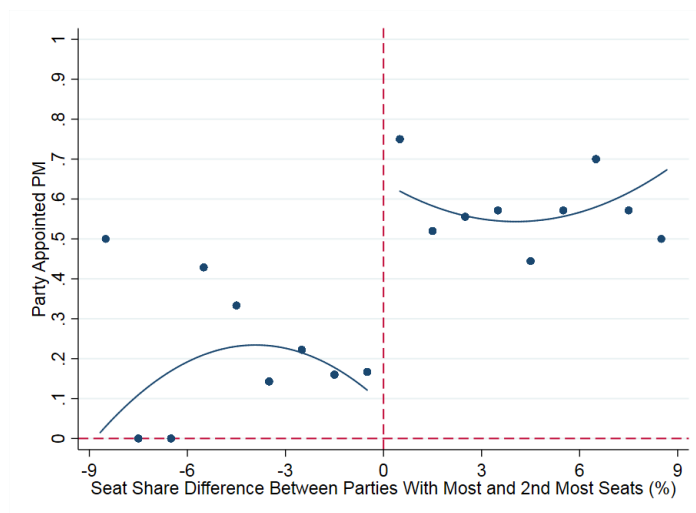
Circles represent estimated effects, using different bandwidth choices (horizontal axis). Whiskers represent the 95% confidence interval based on standard errors clustered at the country level.

Figure A14: Covariate Balance: National Parliaments Data



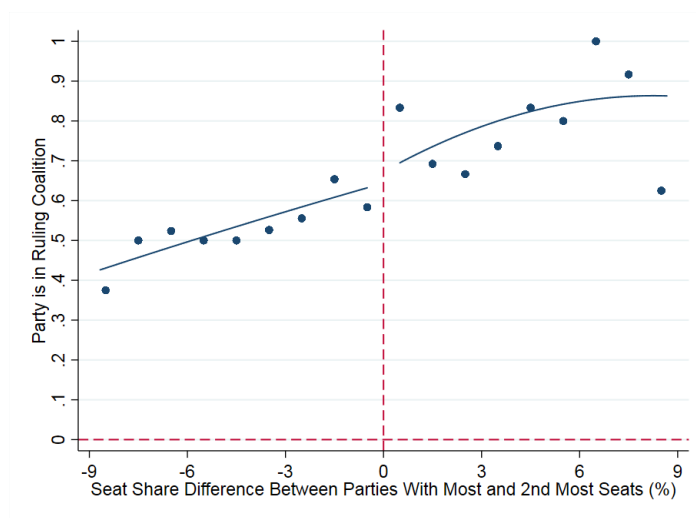
The unit of observation is a country-election-party. Sample is restricted to the two parties with the most seats in the parliament. The running variable (horizontal axis) is the difference in *seat* shares between the two most voted parties: positive for the party with the most seats and negative for the party with second most seats. Circles represent the local averages of the *predicted probability* of the party appointing the prime minister. Predictions are based on regressing a dummy indicating whether the party appointed the prime minister on a set of party family/ideology type (see text for details). Averages are calculated within 1 p.p.-wide bins of seat share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Figure A15: Effect of Having Most Seats in Non-“Dominant” Cases - National Parliaments Data



The unit of observation is a country-election-party. Sample is restricted to the two parties with the most seats in parliaments with non-“dominant” seat compositions (it excludes 51% of elections where the first and third placed parties can form a majority, while the second and third placed cannot). The running variable (horizontal axis) is the difference in seat shares between the two most voted parties: positive for the party with the most seats and negative for the party with second most seats. Circles represent the local averages of a dummy indicating whether the party appoints the prime minister. Averages are calculated within 1 p.p.-wide bins of seat share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Figure A16: Effect of Having Most Seats on Being in Ruling Coalition - National Parliaments Data



The unit of observation is a country-election-party. Sample is restricted to the two parties with the most seats in the parliament. The running variable (horizontal axis) is the difference in seat shares between the two most voted parties: positive for the party with the most seats and negative for the party with second most seats. Circles represent the local averages of a dummy indicating whether the party is part of the ruling coalition (represented in the cabinet). Averages are calculated within 1 p.p.-wide bins of seat share difference (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Table A1: Distribution of Council Sizes

Population	Number of Seats	Number of Municipality-Elections		
		Total	Tie in seats (1st/2nd)	Tie in seats (2nd/3rd)
251 to 1,000	7	15097	822	695
1,001 to 2,000	9	6773	662	214
2,001 to 5,000	11	7064	707	310
5,001 to 10,000	13	3674	365	183
10,001 to 20,000	17	2260	192	89
20,001 to 50,000	21	1369	93	47
50,001 to 100,000	25	469	34	16
100,000+	-	416	23	12

Source: Ley 7/1985, *Reguladora de las bases del régimen local*, article 179.

For municipalities with more than 100,000 inhabitants, one more seat is added for every additional 100,000 inhabitants or fraction thereof, adding one more if needed for odd number of seats.

Table A2: Effect of Being First (Instead of Second) Most Voted:  
Alternative Specifications

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: Main Outcome (mayor serves at least 3/4 of term)</i>					
Party Appointed	0.353	0.185***	0.203***	0.295***	0.241***
Mayor		(0.058)	(0.044)	(0.036)	(0.046)
N		2028	876	5796	5796
<i>Panel B: Outcome is appointing mayor for entire term</i>					
Party Appointed	0.323	0.199***	0.217***	0.306***	0.254***
Mayor		(0.059)	(0.042)	(0.036)	(0.045)
N		1876	876	5796	5796
<i>Panel C: Outcome is appointing mayor for longer than other parties</i>					
Party Appointed	0.374	0.205***	0.221***	0.310***	0.268***
Mayor		(0.061)	(0.045)	(0.037)	(0.047)
N		1998	876	5796	5796
<i>Panel D: Outcome is appointing initial mayor</i>					
Party Appointed	0.360	0.242***	0.249***	0.343***	0.290***
Mayor		(0.062)	(0.045)	(0.037)	(0.047)
N		1892	876	5796	5796
<i>Panel E: Main outcome, sample restricted to cases where "two parties out of top-3 needed for majority"</i>					
Party Appointed	0.370	0.163***	0.200***	0.294***	0.238***
Mayor		(0.062)	(0.047)	(0.038)	(0.048)
N		1898	790	5472	5472
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to the two most voted parties in elections in which they tied in seats. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for a 2nd most voted party that tied with the 1st most voted party (using the specification in column 1). Optimal bandwidths are based on [Imbens and Kalyanaraman \(2012\)](#), being equal to 2.32%, 2.13%, 2.29%, 2.16%, and 2.38% for the five panels, respectively.

Table A3: Effect of Being Most Voted: Cases with a Left-Wing Majority

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: Effect for PSOE (conditional IU being third most voted)</i>					
PSOE Appointed	0.543	0.267*	0.248**	0.417***	0.264*
Mayor		(0.153)	(0.118)	(0.109)	(0.145)
<i>N</i>		155	64	423	423
<i>Panel B: Effect for PP (conditional IU being third most voted)</i>					
PP Appointed	0.102	0.242*	0.231**	0.313***	0.246*
Mayor		(0.146)	(0.110)	(0.101)	(0.139)
<i>N</i>		155	64	423	423
<i>p</i> -value: test of equal effects		0.7826	0.8097	0.1469	0.8412
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to elections in which the two most-voted parties tie in seats and the third-placed party is the *Izquierda Unida (IU)*. Panel A uses only observations regarding the *Partido Socialista Obrero Español (PSOE)* in elections where the *Partido Popular (PP)* is the other “top two” party. Panel B uses only observations regarding the *PP* in elections where the *PSOE* is the other top-two party. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for the 2nd most voted party. The optimal bandwidth is calculated based on the entire sample and is 2.32% (Imbens and Kalyanaraman 2012).



Table A4: Effect of Being Most Voted on Deputy Mayors' Allocation

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: Outcome is share of deputy mayors</i>					
Party Share of Deputy Mayors	0.278	0.094** (0.045)	0.108*** (0.034)	0.183*** (0.029)	0.125*** (0.036)
<i>N</i>		1732	758	4930	4930
<i>Panel B: Outcome is share of deputy mayors in previous period (placebo test)</i>					
Party Share of Deputy Mayors, $t - 1$	0.294	0.007 (0.037)	0.015 (0.035)	0.008 (0.029)	0.026 (0.036)
<i>N</i>		2544	758	4930	4930
<i>Panel C: Outcome is indicator for appointing all deputy mayors</i>					
Party Appointed All Deputy Mayors	0.110	0.079** (0.036)	0.103*** (0.029)	0.151*** (0.026)	0.091*** (0.030)
<i>N</i>		1814	758	4930	4930
<i>Panel D: Outcome is indicator for appointing all deputy mayors in previous period (placebo test)</i>					
Party Appointed All Deputy Mayors, $t - 1$	0.191	-0.009 (0.035)	0.010 (0.030)	-0.003 (0.027)	0.017 (0.034)
<i>N</i>		2444	758	4930	4930
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to the two most voted parties in elections in which they tied in seats. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for a 2nd most voted party that tied with the most voted party (using the specification in column 1). Optimal bandwidths are based on [Imbens and Kalyanaraman \(2012\)](#), being equal to 2.28%, 3.47%, 2.37%, and 3.30% for the four panels, respectively.

Table A5: Effect of Being Most Voted: Heterogeneity by Party Identity

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: Sample restricted to elections where both PP and PSOE are the top two parties</i>					
Party Appointed	0.369	0.166**	0.189***	0.295***	0.235***
Mayor		(0.083)	(0.063)	(0.054)	(0.065)
N		1056	444	3028	3028
<i>Panel B: Sample restricted to elections where either PP or the PSOE are one of the top two parties</i>					
Party Appointed	0.354	0.164*	0.200***	0.277***	0.197***
Mayor		(0.091)	(0.068)	(0.057)	(0.072)
N		808	360	2310	2310
<i>Panel C: Sample restricted to elections where neither the PP or the PSOE are one of the top two parties</i>					
Party Appointed	0.235	0.426**	0.306**	0.461***	0.365*
Mayor		(0.191)	(0.149)	(0.140)	(0.186)
N		164	72	458	458
<i>p-value: test of three effects being equal</i>					
		0.4173	0.7621	0.4655	0.6823
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to the two most voted parties in elections in which they tied in seats. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for a 2nd most voted party that tied with the most voted party (using the specification in column 1). The optimal bandwidth is calculated based on the entire sample and is 2.32% (Imbens and Kalyanaraman 2012).

Table A6: Comparing Magnitude of Effects:  
Effect of Being Most Voted, by Legislature Type

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
<i>Panel A: First and second most voted tied in seats</i>					
Party Appointed	0.353	0.185***	0.203***	0.295***	0.241***
Mayor		(0.059)	(0.044)	(0.037)	(0.046)
N		2028	876	5796	5796
<i>Panel B: Most voted has one more seat than second most voted, but no more “real” bargaining power</i>					
Party Appointed	0.259	0.305***	0.254***	0.431***	0.352***
Mayor		(0.069)	(0.077)	(0.046)	(0.059)
N		1424	252	5862	5862
<i>Panel C: Most voted has one more seat than second most voted and also more “real” bargaining power</i>					
Party Appointed	0.120	0.667***	0.700***	0.618***	0.650***
Mayor		(0.049)	(0.073)	(0.036)	(0.045)
N		1648	160	6382	6382
<i>Panel D: Most voted has a majority of seats</i>					
Party Appointed	0.006	0.978***	0.982***	0.977***	0.976***
Mayor		(0.003)	(0.009)	(0.003)	(0.004)
N		26806	788	56204	56204
Specification:		Linear	Means	Quad.	Cubic
Bandwidth:		Optimal	<1%	Full	Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to the two most voted parties. Each panel focus on a different case of seat composition in the legislature. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for a 2nd most voted party that tied with the most voted party (using the specification in column 1). Optimal bandwidths are based on [Imbens and Kalyanaraman \(2012\)](#), being equal to 2.32%, 4.48%, 7.03%, and 23.18% for the four panels variables, respectively.