Norms in Bargaining: Evidence from Government Formation

Thomas Fujiwara*
Princeton, CIFAR, and NBER

Carlos Sanz*
Bank of Spain

June 2018

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, but consistent with a norm prescribing that "the most voted party should form the government." We first present case studies from several countries and regression discontinuity design-based evidence from 28 national European parliaments. We then focus on 2,898 Spanish municipal elections in which two parties *tie* in the number of seats. We find that the party with slightly more general election votes is substantially more likely to appoint the mayor. Since tied parties should (on average) have equal bargaining power, this identifies the effect of the norm. This effect is comparable to that 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. A model where elections both aggregate information and discipline incumbents can rationalize our results and yields additional predictions we take to the data, such as voters punishing second most voted parties that appoint mayors and municipalities following the norm having less government corruption.

*We thank Albert Falcó-Gimeno, Olle Folke, Nicola Gennaioli, M. Daniele Paserman, Rohini Pande, Francesco Trebbi, Leeat Yariv, three anonymous referees, and seminar participants at Bank of Spain, Brown, Caltech, CEMFI, Chicago Harris, Columbia, Cornell, Cowles-Yale Political Economy Conference, EEA-ESEM Conference, HBS, IDB, IEB Political Economy Workshop, Kellogg, Princeton, Spanish Economic Association Symposium, Stanford GSB, Wharton, and UBC for helpful comments. We are grateful to Albert Solé-Ollé and Pilar Sorribas-Navarro for sharing the corruption data. Vishan Nigam and Susannah Scanlan provided excellent research assistance.

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 determinants of 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).

We start by discussing examples from multiple contexts where this norm's prescription has been explicitly incorporated into the discourse of voters and politicians. Moreover, we leverage a regression discontinuity design (RDD) embedded in government formation episodes in 28 national European parliaments in the postwar period to show that the *party with the most seats is more than twice as likely to form a government* than the party with the second most number of seats.

Next, we then turn our focus to Spanish municipalities, a unique context that allows us to identify this norm with a sharper empirical strategy. Each municipality elects a council by (closed list) proportional representation under the D'Hondt apportionment rule in a single-district election.[1] In its first meeting, the council selects, by *majority* rule, one of its members to hold the powerful position of mayor. 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. Parties A and B thus 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 A appoints the mayor should be equal to B's. Our findings, however, are that party A has a 55% probability of forming the government, while 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.[2] While

[1] Proportional representation allocates seats to parties in proportion to their votes. Exact proportionality is not possible given integer constraints and D'Hondt rule (described in Section 4) allocates seats respecting such constraints.

[2] Only leaders (the higher ranked candidate on the party list) can become mayor. Thus, 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 is that the party with most votes in the *general* election becomes mayor. Section 4 describes these electoral rules in detail.

at first pass it may seem to play an important role, there are several reasons this rule is unlikely to explain our results. 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 voted party and hence the status quo rule cannot fully explain our evidence.[3]

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 variable that fully determines the ranking (votes). This adds to evidence of rank-based decision-making in politics (Anagol and Fujiwara 2016 and Folke, Persson, and Rickne 2016) and is consistent with models where agents make choices overweighting salient attributes (Bordalo, Gennaioli, and Shleifer 2013).[4]

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 (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.[5] 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, standard datasets used in this literature usually contain information only on parties' seat allocations (and not their general election votes).[6]

[3]Sections 4 and 7 discuss additional reasons why the status quo rule cannot explain our results. Note 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 provide evidence that voters punish parties that deviate from the norm and it is unclear why a status quo rule would generate this behavior.

[4]Pope (2009), Kuziemko et al (2014), and Hartzmark (2015) study rank-based decision-making in hospital choice, preferences for redistribution, and investment decisions, respectively. 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.

[5]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).

[6]This is the case of the data from National European parliaments used in this paper (the *European Representative Democracy Data Archive*). 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).

Moreover, in a substantial number of cases, the norm's prescription binds behavior even when it goes against the perhaps most natural factor determining bargaining outcomes: programmatic affinity between parties. In particular, the effect of being the most voted 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.[7]

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. While the results we discussed so far can appear as *prima facie* irrational or puzzling, we present a simple model that can rationalize the results. Elections in our model have two roles: information aggregation and incumbent disciplining. Elections aggregate disperse 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.[8] A norm that matches our results and where voters punish second most voted parties that appoint the mayor, constitutes an equilibrium. The norm also maximizes voters' expected welfare and thus has instrumental value.

We provide three 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. 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. Third, we provide evidence suggesting that the norm indeed affects policymaking and voter welfare: municipalities that follow the norm have fewer instances of government corruption. We also provide evidence suggesting that the norm affects (costly) bargaining delays and government duration in the European parliaments data.

This paper is related to four broad strands of the literature. First, as previously discussed, our

[7] We also document that programmatical 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.

[8] Persson, Roland, and Tabellini (1997) also interpret different equilibria in voting models as different norms.

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 the same issues 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. Existing models suggest that our results have policy consequences. In particular, the norm we study can add first-past-the-post considerations to proportional systems. For example, Lizzeri and Persico (2001) compare how public good provision differs under proportional representation or plurality rule. The key distinction in their theory is that parties under proportional representation maximize vote shares (which translates proportionally to power), while parties under plurality rule with parties maximize the probability of being most voted (a winner-takes-all contest). At first pass, Spanish context fits the proportional representation case. However, our results indicate this characterization misses the key aspect: parties should not simply maximize vote shares but also aim at being the most voted.[9]

In other words, 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. Our results thus also speak to the phenomena where electoral rules that require pluralities are reinterpreted as majorities and its consequences (Maskin and Sen 2016).

Third, our results speak 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 2018) and to the evidence that rank-based decision-making affects political outcomes (Anagol and Fujiwara 2016 and Folke, Persson, and Rickne 2016).[10] 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 variable that fully determines the ranking (votes). This salience of ranks is consistent with theories of "limited attention" and decision-making that over-weights a salient subset of available information (e.g., Bordalo, Gennaioli, and Shleifer 2013).[11]

Fourth, our results bring nuance to the "puzzle of nominal versus real bargaining weights", since variation in seat shares that do not generate "real" changes in bargaining power affects bargaining

[9] The consequences of proportional versus plurality rules are also studied by Persson and Tabellini (2005).

[10] A growing literature focuses on "behavioral" decision-making in political economy contexts. While our results connect to this literature as they suggest a result that might be *prima facie* irrational, the model we propose to rationalize our results is based on agents with standard preferences making rational decision and processing the information available to them fully. Examples of papers in that literature are 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).

[11] Relatedly, our results also speak to why narrow victories go on to be interpreted (or at least marketed) by some as "the will of the people." For example, only 51.9% of votes in the 2016 "Brexit" referendum involved were to leave the European Union and only 50.6% of the votes in the 1995 Quebec independence referendum were against independence from Canada. However, these results went on to be interpreted (at least by some) as "the will of the people."

outcomes[12]

The next section provides examples from multiple contexts on how the norm we study has been incorporated in the political discourse of voters and politicians. It also discusses its implications to broader phenomena, particularly the cases of "pluralities being reinterpreted as majorities" (Maskin and Sen 2016). Section 3 provides the evidence from European national parliaments and Section Section 4 provides the evidence from Spanish local elections. Section 5 describes the theoretical framework and Section 6 provides tests of its additional predictions. Section 7 discusses alternative explanations. Section 8 concludes.

2 Motivating Examples and General Issues

This section discusses examples from multiple contexts where the norm we study has been incorporated into the discourse by voters and candidates. We focus on its puzzling aspects and how it fits in a broader pattern where pluralities are confused with (or reinterpreted as) majorities (Maskin and Sen 2016).

Examples from government formation in Canada, New Zealand, the United Kingdom, Italy, and Spain. The following description of an October 2015 interview with then-sitting Canadian prime minister Stephen Harper illustrates well the issues at play. Harper states that "*my position has always been, if we win the most seats, I will expect to form the government. And if we don't, I won't.* The interviewer presses him by asking "*So, <u>even as the current government, if you're just a couple of seats behind</u>, you wouldn't try to figure out a way to...*" but Harper cuts in to say "*No. [...] I would not serve as Prime Minister [...] we ask people to make a choice of a government, and so I think that the party that wins the most seats should form the government*" (Wells 2015, emphasis ours).

New Zealand's national parliament illustrates a case where this norm has held for two decades and the political costs of deviating from it. Since a 1994 electoral reform, no party obtained the majority of seats needed to form a government. For seven straight elections, the most voted party formed the government. The first deviation from this norm occurred in 2017, when the second-most voted Labour party formed a government. This was perceived as a the decision of a "*maverick*" third-placed party "*kingmaker*" and was met with confusion by voters. As put by a former governor-general (the head of state), "*the leader of the party with the largest number of seats in Parliament has always been able to form a government. While some voters think that will always be the case, it may not.*"[13]

Similar issues are commonly discussed in British politics. During campaigning for the 2015 British general election, the leader of the Liberal-Democrat Party (Nicholas Clegg) stated that "*the party*

[12] 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. A manifestation of this puzzle relates to the previously discussed Gamson's Law.

[13] The New Zealand example is also illustrative since, in the seven elections between 1994 and 2017, two parties (Labour and National) alternated in power. The government formation process in New Zealand is characterized by the *"absence of formal procedures*." The description above and the quotes are from Daly 2017. The reader's comment section in Daly 2017 provides a host of examples of their perception on the matter (e.g., "*clearly the highest polling party has the largest mandate to negotiate with a potential coalition partner*").

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) and Scottish Labour Party leader Jim Murphy categorically stated that "*the biggest party gets to form the government*" (McKinney 2015).

After the March 2018 Italian elections, the Five Star Movement was the most voted party and obtained the most seats in parliament, however falling short of a majority.[14] During negotiations for the government formation, the party argued that the most voted party appointing the minister would be "*more democratic*". As one of the party's candidates put it, *"given that the Five Star Movement received almost a third of all votes and is by far the single most popular political force in Italy, any other choice would be undemocratic*" (The Guardian 2018).

Voters also appear to explicitly agree with these norms. 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 País 2015).[15] Moreover, in multiple instances, 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). These statements also apply to local politics: former PP leader and prime-minister Mariano Rajoy stated that "*we have always supported that the mayor is the person who received the most votes. It follows from a common-sense democratic rule*" (El País 2018). Subsection 4.2 discusses in further detail a related case study (Olivenza's 2011 election).[16]

Majorities versus pluralities. An interesting common aspect in the five contexts discussed above is that they appeal to notions of what is "right," "democratic," or the "people's choice," even though they relate to cases with a plurality, but not a majority, of votes (or seats). A puzzling aspect is that one could also argue that a majority of voters did not vote for the party being labeled "the people's choice." Moreover, the Canadian case also raises another interesting point issue: the sitting prime minister agrees that his incumbency status is less important than "being the most voted" when it comes to having the right to appoint a prime minister.

These examples (as well as our empirical results) illustrate an issue that Maskin and Sen (2016) refer to as the "*serious confusion when a plurality win is marketed as a majority victory*." The concept is that a rule that is inherently majoritarian (parties representing a majority can form a government) is reinterpreted as a pluralitarian procedure (the most voted voted party forms a government). Maskin and Sen (2016) continue to argue that "*understanding of the critical difference between a plurality and a majority could improve politics around the world*" and how this issue has implications to multiple contexts, from the BJP's hindu-nationalism in India to the effects of the negative of the Muslim Brotherhood on Egypt's democratic institutions.

A 2004 campaign speech by PSOE leader José Luis R. Zapatero illustrates this point. Zapatero

[14] The most voted *coalition*, however, was the center-right coalition.

[15] One caveat in interpreting this question is that it does not make it entirely clear what disagreeing with the statement implies. Given the "more democratic" wording and "even without an absolute majority" qualifier, it seems that the logical alternative is that disagreeing implies "it is equally democratic for a first, second, or third most voted party to appoint the form the government if none has a majority," but this alternative is not clearly outlined in the question.

[16] Similar pledges have been made at regional governments, for example in Andalusia (ABC 2004).

promises to attempt to form a government only if his party obtains a ' '*sufficient majority*" but goes on to clarify that "*a sufficient majority is being the most voted party in Spain, it is having more votes than the second place, because this would mean that most Spaniards approve our proposals.*" (El País 2004). The quote conflates majority with plurality in a circular fashion and is striking since his party was then expected to be the "*second place*."[17]

Example of a broader pattern: Donald Trump's nomination. A case that exemplifies how this issue applies to contexts beyond government formation is Trump's nomination as a presidential candidate. The Republican Party's rules clearly state that a *majority* of delegates is needed to nominate a presidential candidate. However, in the early stages of the 2016 primaries, many expected Trump to achieve a plurality, but not a majority, of delegates votes. Commentators (and Trump himself) declared that the candidate with the most delegate votes should be the nominee. Moreover, 62% of Republican voters agreed with the statement that with "*no delegate majority, the GOP nominee should be the one with the most votes*."[18] Moreover, Silver (2016a, 2016b) argued that a reason for Trump's eventual success at securing a majority was that "*Republican voters were swayed by Trump's arguments that the candidate with the most votes and delegates should be the nominee*." If Silver's argument is correct, a rule that was designed to be majoritarian was reinterpreted as voters as pluralitarian, along the lines of Maskin and Sen (2016).

3 Evidence from European National Parliaments

While the previous section discussed examples where the norm we study is incorporated into the political discourse, this section provides evidence from European parliamentary democracies. We present three results. First, most represented parties that have "just one more seat" than second most represented parties are more likely to appoint prime ministers. Second, long (and likely costly) delays in government formation are frequent. Third, cases where the most represented party appoints the prime minister are associated with shorter delays and longer government durations.

Data and context. Our data covers 308 episodes of government formation (following an election) in 28 European countries in he 1944-2010 period.[19] The data contains the party affiliation of the appointed executive (e.g., prime minister) and the number of seats of each party in the lower house of national legislatures (but not its number of votes in the general election).[20] 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*

[17]The context of the speech was indeed a discussion on whether his party (PSOE) would attempt to form a government even if the PP was the most voted (as expected in the polls). Moreover, as a politician engaged in a discussion over government formation, it is unlikely that Zapatero is unaware of the distinction between majority and plurality.

[18]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 occured in April 2016 (Murray 2016 and Flegenheimer 2016).

[19]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, Netherland, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom.

[20]The fact that parties' vote shares in the general election are not collected is illustrative of how, as discussed in the introduction, the previous literature on legislative bargaining largely ignores the votes that generates seat distributions.

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.[21] 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). To the extent that these are not formally coded in laws, they can be understood broadly as norms. Moreover, 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 European parliaments.[22]

3.1 Effect of Having the Most Seats on Appointing Prime Ministers

Empirical strategy. We document this result by leveraging a regression discontinuity design (RDD). A key aspect is that the sample is restricted to i) only include two parties with the most seats in the legislature and ii) exclude cases where one party has a majority of seats. Our final sample includes 504 parties from 252 elections.[23]

Define s_{it} as the seat share of the party with the most seats minus the seat share of the party with the second most seats in country i at election year t. We define the running variable, which varies at the party-country-year level, as follows:

$$x_{pit} = \begin{cases} s_{it} & \text{if } p \text{ has the most seats} \\ -s_{it} & \text{if } p \text{ has the second most seats} \end{cases}$$

where p refers to a political party.

Therefore, if $x_{pit} > 0$, then party p has the most seats (otherwise, has the second most seats). Our outcome of interest is y_{pit}, a dummy indicating whether the party p appointed the relevant executive member for the entire term following the election t. We refer to this outcome as "appointing the prime minister (PM)" throughout. The effect of having the most seats is given by $\lim_{x_{pit}\downarrow 0} \mathrm{E}[y_{pit}|x_{pct}] - \lim_{x_{pit}\uparrow 0} \mathrm{E}[y_{pit}|x_{pit}]$, which can be estimated by a local polynomial regression:

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

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, θ_1 captures the effect of having the most seats (instead of

[21] 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).

[22] More specifically, Diermeier and Merlo (2004) show that the formateur selection is better explained by selection probabilities being proportional to seat shares ("propotional selection") instead of ranks in seats ("selection-in-order"). Proportional selection cannot generate a jump at the RDD cutoff we present in this section. Our results are not inconsistent with Diermeier and Merlo (2004) since: i) they study *formateur* selection, while we study prime minister appointments; ii) their sample contains 11 European countries, as opposed to the 28 in ours; and iii) their maximum likelihood exercise estimates a model that best fits all cases in the data, while we focus on behavior near a RDD cutoff.

[23] The data contains three cases in which the two most represented parties tie in number of 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.

having the second most seats) in a parliament where the two parties with the most seats almost tied in number of seats. Our baseline estimates use a linear specification ($g_0 = g_1 = x_{pit}$), 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 country level.

There are three noteworthy aspects of this setup. First, variables that do not vary across parties within an country-year 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 $x_{pit} = a$ and the other $x_{pit} = -a$). Intuitively, the variation that identifies the results comes from comparing parties within an election.[24] 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 (i.e., will not affect the balance in predetermined covariates around the cutoff). For example, restricting the sample on only the cases where a party does not have a majority of seats cannot create imbalances around the cutoff since "one party having a majority" varies at the election level. 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 country level.

Main result. The graphical representation of our main result is Figure 1a, which plots the probability that a party appoints the PM against the seat share difference between the two parties with the most seats (recall that parties placed third and lower are excluded from the sample). The 504 observations are aggregated into bins of one p.p. width of the running variable (x_{pit}), 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.

The "jump" at the cutoff suggests that a party with "one more seat" is almost 40 p.p. more likely to appoint the PM. Moreover, 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 PM. However, the additional seat that "flips" a party into being the most represented has a sizable impact.

Panel A of Table 1 shows the equivalent regression results. A party with the second most seats that almost ties with the one with most seats has a 20.2% chance of appointing the PM (the "2nd-Place Mean", which is the estimated θ_0 from equation 1). Column (1) indicates that the party with most seats (but almost tied) is 30.3 p.p. more likely to appoint the mayor (the estimated θ_1). Both figures are based on using only the 224 observations within the optimal bandwidth. Column (2) compares the average outcome for parties that are only 1% of seat share apart and finds an even larger effect. Both the linear specifications are robust to the choice of bandwidth (Figure A2). Using the full sample and a quadratic or cubic polynomial yields similar results, shown in columns (3) and (4).

Interestingly, this effect is not driven by an increased probability of being in the ruling coalition (having cabinet positions). Figure A3 shows replicates the exercise of Figure 1a, 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.[25]

[24] Figure A1 presents the distribution of observations, demonstrating this symmetry.

[25] Unfortunately, our data 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.

Covariate balance. 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 PM in the *previous* term against their current seat share difference. The absence of a jump at the cutoff indicates that parties with close seat shares are equally likely to be the incumbent mayor. Panel B of Table 1 present the analogous regression results.[26]

Does one seat make a difference in the ability to form majorities? The effects presented so far can perhaps be explained by the party with most seats having a *numerical* advantage in forming coalitions. For example, it is possible that the first placed party can form a majority with only the third-placed party, while the second placed cannot. To investigate if such differences drive our results, as opposed to a norm of the most represented party forming government, Figure A5 repeats the exercise of Figure 1a, but using a subsample of 254 observations that excludes all cases where the party with the most seats could form a majority coalition with the third most represented party, while the second most represented cannot.[27] The effects are strikingly similar to those from the main sample. This is consistent with the effect of having most seats being driven by a norm, as opposed to higher bargaining power associated with the ability to form different coalitions.

However, even in the parliaments in Figure A5's sample, it is still possible that the party with most seats has higher ability to form majorities 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, such subsamples would include only a small number of observations. This highlights the value of focusing on Spanish municipal governments in the next section: its data provides 2,898 cases where the two most voted parties have the exactly same number of seats, and hence its ability to form coalitions is the exact same.

3.2 Does the Norm affect Bargaining Delays and Government Duration?

Delays in government formation. Delays in government formation can be costly. For example, it took 194 days for the representatives elected in the June 2007 Belgian election to form a government. In these six months, a government with caretaker status was "*unable to take policy decisions*" and a crisis ensued. A similar case happened after the 2006 Czech Republic election, which led to a 7-month period without a government and left "*legislation and important reforms in a state of limbo*." Previous literature associates delays in government formation with financial market volatility.[28]

Moreover, such delays are also frequent. In our sample of 252 European elections where no party obtained a majority of seats, it takes on average 38 days for a government to be formed (the median delay is 33 days). In 19% of cases, the delay is longer than 60 days. Delays are longer when the most represented party does not have clear dominance. Excluding the cases where the most represented party can form a majority coalition with the third most represented party (while the second most

[26] Party identity is also balanced around the RDD cutoff. Andersson and Ersson (2014) classify the parties in our data into 12 ideological families (e.g., social-democratic, liberal). We regress our outcome on a set of dummies indicating each family. We use the predicted value from this regression as an outcome (Panel C of Table 1 and Figure A4). If a particular "family" was more likely to be on a particular side of the cutoff, we would observe a non-zero effect.

[27] Laver and Benoit (2015) classify these exact cases as "legislatures with a dominant party."

[28] The Belgian and Czech examples and quotes are from Golder (2010), which also surveys previous literature on the causes and consequences of bargaining delays in government formation.

represented cannot) increases the average delay in the sample to 50 days and the share of cases with delays longer than 60 days to 32%. Figure A6 plots the cumulative distribution function of delays in government formation in our data for each of the subcases discussed.

The party with most seats appointing the PM is associated with shorter delays. Panel A of Table 2 presents results from regressing the (log) number of days taken to form a government against a dummy indicating whether the party with the most seats appointed the PM. The result in column (1) indicates that, when the PM is from the most represented party, the time taken to form a government is almost 35% shorter. This result is robust to controlling for country fixed effects (column 2). One possible confounding factor is that cases where the most represented party is stronger (e.g., has a majority or a large plurality of seats) are associated with both shorter delays and the eventual appointment. To address this issue, column (3) controls for a set of dummies capturing Laver and Benoit's (2015) five classes of the implied bargaining power given the seat distribution. The coefficient remains negative and sizable. Lastly, column (4) adds to this specification a set of dummies capturing the "ideological family" (e.g., conservative, liberal, green) of the most represented party in the sample.

The results suggest that the norm we study may have consequential welfare, as they reduce (costly) delays in government formation. Furthermore, another (more speculative) interpretation is that these norms arise exactly to shorten such delays. It is important to caveat, however, that the results in Table 2 are based on correlations and, although robust to many controls, it is difficult to fully rule out the possibility of other confounding factors.

Term lengths. Panel B of Table 2 repeats the exercise described above using the length of the government (the number of days it remained in power) as the outcome of interest.[29] If indeed a norm prescribing that a PM from the party with the most seats is a "more democratic" choice that better reflects "the will of the people," it is possible that governments formed by such parties should have an advantage in staying in power too. This is what the results suggest, as PMs from parties with the most seats stay in power 18-21% longer. These results are also robust to a set of controls, but the caveats regarding the result being based in correlations also apply.

4 Evidence from Spanish Municipal Governments

As discussed in Subsection 3.1, while the results from European national parliaments indicate the existence of the norm we study, they have the drawbacks of a small sample size and the inability to fully control for potential differences in bargaining power associated with additional seats. This motivates this section's focus on Spanish municipalities. Given Spain's numerous municipalities (and 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."[30]

[29] We approximate term length by the days between the election that led to that government and the next election.

[30] 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 that 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).

4.1 Context, Data, and Empirical Strategy

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.[31] Political parties must also obtain at least 5% of the votes to receive seats.[32]

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 7 discusses why this status-quo rule is unlikely to drive our results.[33]

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 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 "off-schedule" general elections:

[31] This rule is also known as Jefferson rule and used to assign US House districts to American states according do their population. It works as follows. The total votes cast for each party in the electoral district is divided, first by one, then by two, then three, up to the total number of seats to be allocated. Let p be the number of parties and s the number of seats. Then a grid of numbers with p rows and s columns is created, where the entry in the ith row and jth column is the number of votes won by the ith party, divided by j. The s winning entries are the s highest numbers in the whole grid (each party is given as many seats as there are winning entries in its row).

[32] This system refers to municipalities with more than 250 inhabitants. Other municipalities use a different (open-list) electoral system and are excluded from our sample. Sanz (2017) and Gonzalez-Eiras and Sanz (2018) estimate the effects of the change in electoral rules at the 250-inhabitant threshold on voter turnout and women's representation, respectively.

[33] 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.

citizens only vote in municipal elections every four years.[34]

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 taxes on residential properties, businesses, vehicles, and collect fees and user charges. Total tax and fee collection by municipal governments is approximately 4% of the Spanish GDP. Hence, municipal-level policymaking is consequential to voters.[35]

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 municipalities. 2,898 elections have the first and second most voted parties tied in number of seats.[36]

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.[37] 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

[34] If a mayor resigns, is convicted of of a crime, or dies, a new mayor is chosen by the same procedure.

[35] 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 choose to provide other services (e.g., childcare).

[36] 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. Including the 1979 and 2015 data in our regressions does not affect our results (Table A3).

[37] We do observe the party affiliation of deputy mayors, as described in Subsection 4.2.

the term in power during the term belongs to that party.[38]

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."

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%).

Empirical Strategy. We implement a regression discontinuity design (RDD) with the key aspect 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 5,796 observations (from 2,898 elections).

Conditional on this sample definition, the implementation of the RDD-based empirical strategy follows closely the one described in Subsection 3.1. Specifically, we estimate equation (1), but now the running variable x_{pit} for party p in municipality i at time t is defined as follows:

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

where v_{it} is the vote share difference between the first and second most voted parties. Thus, if $x_{pit} > 0$, party p has the most votes ("first-place") and it has the second most votes otherwise. As discussed in Subsection 3.1, this implies that 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 be mechanically be distributed symmetrically around the RDD threshold.[39] It also implies that conditioning which observations enter the sample by a variable that varies only at the election or municipality level will not affect the internal validity and balance of the RDD design. 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.

[38] 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 sensitive 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).

[39] Figure A7 presents the distribution of observations, demonstrating this symmetry.

4.2 Identifying the Norm: Effects of Being Most Voted

The graphical representation of our main result is Figure 2a, which plots the probability that a party appoints the mayor against the vote share difference between the first and second most voted parties (parties placed third and lower are excluded from the sample). The 5,796 observations are aggregated into bins of 1 p.p. width of the running variable 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.

A clear jump at the cutoff is visible in the Figure 2a (the graphical counterpart to θ_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.[40]

A perhaps surprising pattern in Figure 2a 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.[41]

Panel A of Table 3 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 θ_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 θ_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). Results are robust to the choice of bandwidth (Figure A10). 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%.

Table A2 replicates Panel A of Table 3 when defining "appointing the mayor" in different manners. Panel E of Table A2 restricts the sample to the cases where two of the top three most voted parties are needed to form a majority. Table A3 repeats this analysis incorporating the data for 1979 and (when possible) 2015. In all cases, the effects are similar in magnitude to those in Panel A of Table 3.

Interpretation. As discussed in the introduction, comparing the two most voted parties that obtained almost the same number of votes allows us to identify a norm. When comparing two groups that should have, given formal rules, the same bargaining power, finding differences in bargaining

[40] In 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.

[41] 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 A8 replicates Figure 2a for specific council sizes and indicates the lack of slope is not driven by such compositional effect. Similarly, Figure A9 replicates Figure 1a for specific council seat compositions.

outcomes isolate the effect of the norm. Note that an effect size of 20 p.p. is both consistent with 20% of the municipalities having a norm that always binds and with all municipalities in the sample having a norm that only binds with 20% probability (or a combination of these two extreme cases). Our research design does not allow to pin down which is the case.

Covariate Balance. To assess the validity of the RDD, Figure 2b repeats the exact same exercise from Figure 2a, 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 A11 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 3 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 parties running, municipal income, 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.[42]

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 and 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.[43]

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 A4 provides the corresponding table.[44]

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

[42] For every party with forcing variable $x_{piit} = a$, there is also a party with $x_{pit} = -a$ 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 A7 which shows how many observations are in each bin of Figure 2.

[43] 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 only did so once at the regional level, under unusual circumstances in the Basque Country.

[44] The estimates on Table A4 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.

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.

An example. 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 is interpretation is unlikely. First, as previously discussed, 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 6. 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 A8, however, indicates that RDD "jumps" are larger in municipalities with larger councils, which are also those with more inhabitants (Table A1).[45] Fifth, it does not appear that mayors share power with other parties. Appendix A (and Figure A19 and Table A7) discusses how this can be inferred from the appointment of deputy mayors (*tenientes de alcalde*), the second most visible position in municipal government.

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 A12 replicates Figure 2a, but separately plotting the cases where i) both the PP and PSOE are the two most voted parties, ii) only one of them was amongst the two most voted parties, and iii) neither are amongst the two most voted parties. 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

[45] 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.

third placed) party supports the first most voted.

Is the most voted effect different for powerful parties? Figure A12 suggests that the norm has the same effect on a party, regardless of whether it is one of the two dominant parties is Spanish politics or not. To further probe whether the norm has differential effects given how powerful a party is, Figure A13 repeats the exercise of Figure 2a plotting separately the cases where the party is in power (i.e., holding the main executive position) at the national (or, alternatively, regional) government at the time of the municipal election. The effects are similar regardless of whether the party is in power or not. We return to this issue on Section 5, in light of our theoretical framework.

4.3 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.[46]

We can thus repeat our RDD, with the running variable also redefined accordingly (the difference between second and third most voted). Figure 3a provides a graphical exercise similar to Figure 2a 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 2a. Note that no discontinuity is visible in the placebo graph (Figure 3b) which plots lagged outcomes—close second and third most voted parties are equally likely to have appointed the incumbent mayor. Figure 3b also increases confidence that the jump observed in Figure 3a is not driven by noise.

Table 4 provides the regression results, following the template of Table 3. 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 A14 provide the evidence of covariate balance (as expected in a RDD). Figure A7 shows the number of observations in each bin of Figure 3a. Figure A10 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.[47] Moreover, it suggests the similar effects of being

[46] Our sample contains 1,565 elections (from 1,204 unique municipalities) satisfying these two conditions.

[47] Spanish electoral rules do not create any differential treatment of second instead of third most voted parties.

most voted are also not driven by the status quo rule. Section 7 further discusses this issue.

Finally, a similarly defined third (versus fourth) most voted effect and found effects is (statistically) close to zero. This is perhaps expected, since such lower ranked parties rarely appoint the mayor.[48]

4.4 Comparison to 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.[49] 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.[50]

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 other parties, as any two parties can form a majority.[51]

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.[52] Figure 4 provides the regression discontinuity plot these different subsamples. The blue

[48] Figure A15 presents a graphical analysis similar to Figures 1a and 2a, redefining the sample accordingly. Note the y-axis range matches that of Figure 2a for comparison. This 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

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 comparison between the "blue circles" and "red triangles" is a perhaps surprising result. It suggests that the effect of "being labeled the most voted" is of similar magnitude as "being labeled the most voted and obtaining an extra seat." The corresponding estimates are provided on Panel B of Table A5, which show that the estimated effect is larger, but not substantially so. Based on our baseline specification (column 1), the effect of being labeled the most voted is 60% the size of the effect of being labeled the most voted and obtaining an extra seat. Since the effect of additional seats have been the focus of previous literature (e.g., Gamson's Law) but studying the effect of rank labels is a more novel aspect of our study, it highlights the importance of the norm in driving overall bargaining outcomes.

Similarly, 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. Panel C of Table A5 provide the corresponding estimates. The specification on column (1) indicates that the effect of simply being labeled the most voted is 28% the size of the effect of the being labeled the most voted and obtaining an additional seat that awards real bargaining power. This again 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 (e.g., Warwick and Druckman 2001, Frechette, Kagel, and Morelli 2005) has noted that nominal bargaining affecting outcomes conditional on real bargaining constitutes a puzzle. The existence of the norm we study adds an additional mechanism that may help explain it.[53]

Lastly and as expected, receiving a majority of seats (yellow diamonds) makes a party almost surely appoint the mayor. Panels D of Table A5 provides the corresponding estimates.

5 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 empirical analysis. 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

[53] A caveat when comparing the effects in different panels of Table A5 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 49.

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.

The model illustrates how a norm that might appear, *prima facie*, to be irrational behavior can be sustained in equilibrium by a set of rational agents with standard preferences. Additionally, this equilibrium maximizes voters' expected welfare. Although our analysis remains agnostic on the issue of equilibrium (norm) selection, this sheds light on why such a norm might arise in multiple contexts. Moreover, the model setup is relatively general and its main results are not based on a specific bargaining procedure or electoral rule (such as proportional representation). The key aspect of the model is the "parliamentary" form of government: elected representatives appoint the executive and voters' and representatives' preferences over this appointment can differ.

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).[54] 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.[55] 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.[56]

Parties' preferences and the bargaining procedure are thus independent of voters' welfare and the 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, but the choice of mayor may be determined by factors that are orthogonal to their interest. If states

[54] 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 C discusses how this can be relaxed.

[55] 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.

[56] 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.

were directly observable, voters would always award 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 $\boldsymbol{p}_t$ denoting the vector $[p_t^A, p_t^B, p_t^C]$. Moreover, voters and parties face uncertainty about the vector $\boldsymbol{p}_t$, which is drawn every period from a (common knowledge) distribution $G(\boldsymbol{p})$ that is serially uncorrelated and identically distributed over time. Each voter individually observes a signal σ_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.[57]

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 $G(\boldsymbol{p}_t)$. 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 $\boldsymbol{p}_{t+1}$ and state s_{t+1} will be drawn, making previous information irrelevant.

We assume that $G(\boldsymbol{p}_t)$ 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 B provides an example of a $G(\boldsymbol{p}_t)$ 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 $\boldsymbol{p}_t$. 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 *i*ncumbent parties A^I, B^I, or C^I that were in office in the preceding period or voting for one of *ch*allenger parties A^{Ch}, B^{Ch}, or C^{Ch}. In other words, for each of the three party types, 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 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 multiple leftist parties (some regional or municipal) that can replace

[57] All realizations of $G(\boldsymbol{p})$ satisfy $p^A + p^B + p^C = 1$.

one another.[58] While the model assumes a specific bargaining procedure, Appendix C discusses how the results are robust to allowing multiple rounds of bargaining and rents to be divisible across parties. Note also that the model does not rely on a specific electoral rule (e.g., proportional representation), with the key aspect being that voters do not directly elect the executive.

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.[59] We highlight that the information structure is serially uncorrelated: a new independent draw of $\boldsymbol{p}_t$ 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 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). This interpretation of multiple equilibria as different norms in 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.

Appendix B presents the proofs. 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 Sections 3 and 4. Those results, however, can be captured by the following equilibrium, where the most voted party appoints the mayor in every period.

Proposition 2. If $G(\boldsymbol{p}_t)$ 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 the most 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,

[58] 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.

[59] 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.

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.

The 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. The equilibrium in Proposition 2 generates the RDD "jump" we study in Sections 3 and 4. 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 equilibria as norms: the behavior of voters and parties is mutually self-enforcing.

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 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. 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 2016), consumers (Pope 2009), and investors (Hartzmark 2015).[60]

Note also that our model predicts that the norm applies uniformly to all parties. This is consistent with the evidence we discuss in Subsection 4.2: the effect of the norm is similar regardless of whether the most voted party is particularly powerful at higher levels of government or aligned with the third-placed.

Welfare. The equilibrium in Proposition 2 maximizes expected voter welfare. It guarantees the party most likely to match the state (given *aggregate* information) of the world appoints the mayor in every period. This is not the case in the equilibrium described in Proposition 1. While the model not directly address equilibrium selection (why the the norm is adopted or not), the fact it is optimal for voters can provide an explanation for its prevalence.[61]

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

[60] 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.

[61] Proposition 2, however, specifies the necessary conditions for the equilibrium with the norm we study, and hence clarifies why it might not occur in some contexts. It is possible to construct an equilibrium where the second (or third) most voted must appoint the mayor. However, voters' expected welfare would be lower in such equilibrium.

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 4 pin down the share of municipalities that follow the norm.[62]

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 and motivates the evidence we provide in the next section.

Proposition 2 provides conditions for the norm to occur: for any party, the benefit of deviating from the norm is smaller than the benefit of continue to follow it and perhaps obtain rents in a future period when it is the most voted party. Whether this condition is satisfied is not directly observable in the data. However, it can be approximated empirically in two manners. First, one can observe if the third-placed party actually has been (or will be) the most voted in a past (or future) election. Second, the condition is more likely to be met when the third most voted party has a larger vote share.[63] This suggests a test based on the heterogeneity of effects we present in the next section.

Lastly, the model model predicts that municipalities with the norm have mayors of better "quality" and thus higher welfare, which we test by checking whether municipalities that follow the norm have fewer instances of corruption.

6 Empirical Implications of the Theoretical Framework

The previous subsection outlined three empirical predictions of the model that we test in this section.

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

This subsection provides evidence suggesting that voters enforce the norm we study. While Subsection 2 discussed that voters and politicians in multiple contexts appear to explicitly agree with the norm prescriptions, here 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 placed 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 4).[64] 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.

[62] Voters also prefer the second most voted party to appoint the mayor instead of the third most voted and thus 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 do so in a period). In such cases, norms where voters impose that the second most voted take preference over the third most voted can also be an equilibrium.

[63] Intuitively, if vote shares are 45.5%, 44.5%, and 10%, the third party is unlikely to be the most voted in future elections. Compare that to an election where the vote shares are 35.5%, 34.5%, and 30%, where it is more likely that the third-placed party will be the most voted in a future election.

[64] 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.

We estimate a triple-differences equation for vote share (v) of party p in municipality i at year t:

$$(v_{pi,t+1} - v_{pi,t}) = \alpha + \beta m_{pit} * f_{pit} + \gamma m_{pit} + \delta f_{pit} + \epsilon_{pit}, \tag{2}$$

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, γ 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 parties that appoint mayors are rewarded compared to second most voted parties that do the same. Note that equation (2) nets out municipality-party fixed effects and time effects are absorbed into α. Standard errors are clustered at the municipality level.

The results are shown in Column (1) of Table 5, 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 indicators 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 that 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:

$$(v_{pi,t+k} - v_{pi,t}) = \alpha_k + \beta_k m_{pit} * f_{pit} + \gamma_k m_{pit} + \delta_k f_{pit} + \epsilon_{pit} \tag{3}$$

separately for k equal to -3, -2, -1, 1, 2, and 3. Figure 5b plots the β_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).[65] 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 most voted parties after their appointment.

While Figure 5b indicates a non-zero effect four years after the deviation from the norm, it is not

[65] This implies we estimate $(v_{pi,t+k} - v_{pi,t}) = \alpha_k + \beta_k m_{pit} + \epsilon_{pit}$ when using only the second most voted parties and $(v_{pi,t+k} - v_{pi,t}) = (\alpha_k + \delta_k) + (\beta_k + \gamma_k) m_{pit} + \epsilon_{pit}$ when using only the first most voted parties.

as conclusive regarding whether parties continue to be punished eight or twelve years later. The point estimates for are positive but not significantly different from zero, so it is not possible to discern if the punishment fully dissipates after two elections, or if it only dissipates partially and we lack the statistical power to detect the smaller effects in later elections.

6.2 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 6 provides evidence supporting this prediction. It repeats our estimation of the main results (Table 3), 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.[66] The effects are substantially larger in Panel A, and it is possible to reject that the the effects in both subsamples are the same in all specifications across columns (at the 5% level). Figure A16 provides the graphical counterpart.[67]

Table A6 repeats this exercise using another criterion to separate the two subsamples. Panel A focuses on the cases where the third most voted party has been (or will be) the most voted party in at least one of the last (next) three elections. Panel B focuses on the remaining cases. Again, the effect is substantially larger in Panel A. Figure A17 provides the graphical counterpart.[68]

Lastly, we test whether voters' punishment for deviations from the norm is more evident in elections in cases 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). Column (4) of Table 5 estimates equation (2) for the subsample in the data where the vote share of the third-placed party is above the sample median, while column (5) repeats this for the case below the sample median. The estimated punishment for deviations from the norm (β) is larger in the former case.

6.3 Is Following the Norm Associated with Less Corruption?

The model predicts that municipalities under the equilibrium with the "most voted appoints the mayor" norm have mayors that voters prefer and higher welfare. This section presents results suggesting that municipalities following the norm have lower political corruption. Given that it is not possible to directly observe voter welfare or easily infer the quality of policymaking from observable policies (e.g., budgets and spending) or outcomes that governments have little control over (e.g., income or population growth), the focus on corruption is well suited for our purposes. Government corruption in

[66] We use the median of the sample with optimal bandwidth, which is a vote share of 16.5%. In the above (below) median subsample, the average vote share of the two most voted parties is 40% (33%) each, with the third most voted obtaining 12% (22%).

[67] Figure A16 is constructed similarly to Figure 1a, but for each subsample. Figure A16b illustrates why the effect for the below median subsample varies across columns in Table 6, there is a 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 A9 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).

[68] This criterion is the closest empirical approximation of the conditions on Proposition 2 (that the probability the third most-voted goes is a winner in the future is sufficiently high). Compared to the criterion of Table 6, it has the drawbacks that it requires a reduced sample (it can only be defined for cases where we observe the outcomes of the last and next three elections) and that most of the observations appear on Panel B.

Spain is usually linked to *municipal* regulation of land use (in a typical case, local officials take bribes in exchange for amendments to land use plans and building permits). Moreover, corruption is relevant to voters and perceived as costly, both in Spain and other contexts (Fisman and Golden 2017).

We use a measure of corruption based on newspaper reports from Solé-Ollé and Sorribas-Navarro (forthcoming) that covers all Spanish municipalities in the 1991-2015 period. It contains a dummy ($corruption_{it}$) indicating whether a corruption case was uncovered in municipality i during the electoral term starting at year t. A corruption case occurs in 5.7% of the observations in our sample, and 22.5% of municipalities experienced at least one corruption case during 1991-2015, indicating that corruption is widespread.[69]

Whether a municipality is following the norm we study is not directly observable, but we can construct a proxy variable ($norm_i$) in the following manner. For each municipality in the sample, we calculate the number of times in which the two most voted parties tied in seats and their vote shares were less than 1 p.p. apart. If this number is zero, we code $norm_i$ as missing, otherwise, we code $norm_i$ as the share of times after such close elections that the most voted party appointed the mayor. The intuition, based on our previous analysis, is that a strong signal of whether or not a municipality follows the norm is only available when observing parties almost tying in the number of votes.

To estimate the effect of the norm on corruption, we regress $corruption_{it}$ on $norm_i$, and study the robustness of the results to a variety of fixed effects and controls. While the outcome varies at the municipality-electoral term level, $norm_i$ does not vary across time within a municipality.[70]

Column (1) in Table 7 presents the estimate from this regression. To interpret the magnitudes, consider a municipality following the equilibrium without the norm. It would have a value of $norm_i$ of approximately 1/2 and a 5.7% probability of observing a corruption during a 4-year electoral term. A municipality following the norm in all periods (as in the equilibrium in Proposition 2) would have $norm_i = 1$ and a 4.4% probability of observing a corruption scandal. This difference is not only statistically significant but also economically substantial. Columns (2)-(5) sequentially add a number of controls (described in the Table 7's notes). We highlight that including province-year effects does not affect our results. Intuitively, the results hold when comparing two municipalities from the same province in the same year, which is remarkable since most confounding effects are likely to systematically vary at this level. However, it is also important to caveat that the results in Table 7 are based on correlations and, although they appear robust to controlling for many relevant factors, it is difficult to fully rule out the possibility of other confounding factors.[71]

[69] Our corruption data was originally created by Fundación Alternativas, which hired a journalist in each Spanish province to compile news items referring to its municipal corruption scandals by looking at the municipal, regional, and national press. Solé-Ollé and Sorribas-Navarro (forthcoming) expanded the dataset for the 1999-2007 period by conducting internet searches of newspapers. The same authors have expanded the data to the 2007-2015 period by searching all national and many regional newspapers in the Factiva archive, relying on a machine learning algorithm to identify the actual cases of corruption occurring by municipality. The data is harmonized to be comparable throughout the 1991-2015 period. Further informaton on the data can be found in Fundación Alternativas (2007) and Solé-Ollé and Sorribas-Navarro (forthcoming), which also present a general overview of municipal corruption cases in Spain.

[70] This implies that we cannot include municipality fixed effects. The sample is an unbalanced panel of 2,390 observations from 450 unique municipalities. $norm_i$ is equal to zero 44% of the times and equal to one 52% of the cases (with the few remaining cases being equal to 1/2 or 2/3). Note that the fact that we only observe one (or a few) close elections for each municipality only generates classical measurement error in $norm_i$ and thus can only generate attenuation bias.

[71] The results are very similar if we use a probit instead of a linear probability model.

7 Alternative Explanations

Status quo rule. The only differential *institutional* treatment of parties by rank of their votes in Spanish municipal elections is the status quo rule described in Subsection 4.1. If no candidate receives a majority of votes in the council election, the party with the most votes appoints the mayor. While, at first pass, this appears to be likely to explain our results, we believe the status-quo rule cannot be the main driver of our results. 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 4.3 and thus cannot account for the entirety of our evidence. Appendix D discusses in further detail other reasons why we believe the status quo rule does not play a role in explaining our results.

Politicians' outcome bias. A possible alternative explanation is that politicians that receive the most votes become inherently more motivated and exert more effort into forming a government. We believe there are two reasons why this is unlikely to explain the entirety of our results. First, we estimate sizable effects, which would suggest motivation has perhaps unreasonably large impacts on political outcomes. Second, it is not clear how politician motivation can generate other results such as voters punishing those that deviate from the norm and the heterogeneity by strength of third place.

Voters prefer "winners." Another explanation for our results relies on assuming that voters' preferences are such that they inherently prefer the most voted party to appoint mayor. This could be because voters also present the outcome bias described above or because "winning" creates symbolic value to a party. This would in turn imply voters punish mayors from second most voted parties in future elections, tilting bargaining outcomes towards the most voted party.

First, note that this "alternative" explanation is in many aspects similar in spirit to our model. One of its goals is illustrating an information aggregation rationale for voters to prefer, in equilibrium, the most voted party to appoint the mayor. This alternative explanation, on the other hand, relies on directly assuming that voters prefer mayors that are the most voted. We do not mean to imply, however, that the mechanism proposed in our model is more appropriate simply because it does not rely on direct assumptions about preferences. On the contrary, we aim to bring attention to the complementary character of what may appear at first pass as distinct mechanisms. Perhaps because there is an equilibrium logic for preferring the most voted parties to appoint mayors, voters incorporate this into their preferences over time.[72]

Given that our model provides a rationale for voters to prefer most voted parties to appoint mayors, many of its empirical predictions would be the same in other explanations based on voters' inherently having such preferences. We highlight two predictions, however, that are obtained naturally from our model but require further assumptions to be explained by this alternative explanation. First, it is straightforward to extend the logic of our model to generate the second-versus-third most voted results discussed in Subsection 4.3. In the alternative explanation, one would need to assume that voters not only prefer election "winners" over "losers," but also "runner-ups" over "third-places."

[72] Note that, in our model, the concept of social norms is understood as different equilibria. In this alternative explanation, social norms would be encoded in voters' preferences. Both equilibrium selection and formation of preferences are not straightforward issues to be modeled and, as suggested above, they might be complementary processes.

While this may appear natural, recall that these are proportional representation elections, where the "runner-up" label should be, given the electoral rules, meaningless.

Second, our model predicts that the effects are stronger when the third-placed party has a higher vote share. The same prediction is not entirely clear in the alternative explanation. One could envision that the punishment for "second most voted mayors" has larger effects on bargaining outcome when the third-placed party is weaker (since this would imply that the two most voted parties are closer to winning a majority in the future, so the punishment in future elections is more likely to matter). If that is the case, this alternative explanation predicts the opposite of the results in the previous subsection. On the other hand, one could also argue that competition among parties is stronger when the third party has a higher vote share, yielding the opposite prediction.

The central government prefers winners. Another mechanism in the similar spirit would be based on assuming that the central government prefers most voted mayors (e.g., it is more likely to award them funds). As before, it is not clear why a rational and fully informed central government would have such preferences and the argument requires that the central government presents some outcome bias or an inherent preference for "winners" (and also for second most voted over third most voted parties). Similarly to the discussion above, the issues related to "assuming an agent prefers the most voted to appoint mayors" also apply here.

We highlight, however, three pieces of evidence that are easier to reconcile with our model than with an explanation based on the central government preferring most voted mayors. First, it is not clear why the heterogeneity by strength of the third-place would arise under this alternative explanation. Second, as discussed previously, the concept of "it is more democratic to appoint the most voted" appears to be incorporated into the opinions of voters and the discourse of local and national politicians. If the advantage of most voted parties was entirely due to instrumental reasons related to the ability of obtaining more resources from the central government, it is not clear why that would be the case (especially when it comes for national politics, which is the highest level of government itself). Third, the case studies and cross-country evidence discussed in Section 2 relates to national governments too.

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 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 most voted should be stronger in municipalities with more frequent cases of ties in number of seats.[73] Panel A of Figure A18 repeats our main RDD plot (Figure 2a) 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.[74] 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.

Lastly, note that discussion above regards parties making agreements at the municipal level. In principle, one could envision parties would making a nationwide (or province-wide) agreement and forcing its local politicians to follow it. Many of the issues raised above would also apply to this explanation. For example, it is not clear, why would parties find this particular agreement useful (versus any other that split mayoral allocations evenly across first and second most voted) or why voters would punish parties that deviate from the norm. Moreover, there are two pieces of evidence pointing out that the norm we study operates at a local level. First, Figure A12 indicates that our main result is of similar magnitude regardless of whether the two most voted parties are (both, one of, or neither) the two larger national parties (PP and PSOE) or not. Presumably, incentives for a large-scale agreement would not be the same for the parties with nationwide coverage and other parties. Second, the punishment of second most voted parties that appoint mayors occur at the municipal level (when a party in a given municipality deviates from the norm). It is not clear why an agreement at the national or regional level would lead voters to punish parties for local deviations.

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 is consistent with the existence of a norm that voters enforce by punishing parties that deviate from it. This can explain why parties follow the norm even when there is a natural alternative, such as parties that are ideologically close forming a winning coalition.

While we provide evidence suggesting countries following the norm have shorter delays in government formation (and governments that last longer) and Spanish municipalities following the norm have lower corruption, we believe that further studying the welfare consequences of the norm would be a fruitful avenue for other research. Lastly, another interesting aspect to be studied is why some jurisdictions "adopt" this norm or not. In light of our model, that would require understanding the equilibrium (norm) selection mechanisms.

[73] Presumably, previous experience with costly bargaining would be the reason for norm to arise in future cases.

[74] 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.

References

ABC (2004). Chaves acepta el reto de Teófila y promete no gobernar si no es el más 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/*$_1$*83979.html*.

Agranov, M., J. K. Goeree, J. Romero, and L. Yariv (2018). What makes voters turn out: The effects of polls and beliefs. *Journal of the European Economic Association 16* (3), 825–856.

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. *American Economic Review 102* (1), 576–593.

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.

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. *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. *American Economic Review 99*(4), 1145.

Daly, M. (2017). National could get many more votes than any other party and still not be in the next government. *Stuff, https://www.stuff.co.nz/national/politics/97079414/national-could-get-many-more-votes-than-any-other-party-and-still-not-be-in-the-next-government*.

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.

El País (2004). Zapatero anuncia que solo intentará formar gobierno si logra más votos que Rajoy. *https://elpais.com/diario/2004/01/19/espana/1074466803*$_8$50215*.html*.

El País (2015). Clima social tras las elecciones del 24-M. *http://elpais.com/elpais/2015/06/19/media/1434733713*$_9$25358*.html*.

El País (2018). Rajoy critica a los ayuntamientos del cambio por repartirse despachos para arrinconar al PP. *https://politica.elpais.com/politica/2018/05/05/actualidad/1525526954*$_7$12707*.html*.

Europa Press (2007). Rajoy emula a Zapatero y se compromete a no gobernar si su partido no es el más 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 alcaldía de Olivenza, pueblo natal de Vara, dado que IU no apoyará a ningún 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.

Fisman, R. and M. A. Golden (2017). *Corruption: What Everyone Needs to Know*. Oxford University Press.

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?*$_r = 0$.

Folke, O., T. Persson, and J. Rickne (2016). The primary effect: Preference votes and political promotions. *American Political Science Review 110*(3), 559–578.

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.

Fundación Alternativas (2007). Urbanismo y democracia. Alternativas para evitar la corrupción.

Golder, S. N. (2010). Bargaining delays in the government formation process. *Comparative Political Studies 43*(1), 3–32.

Gonzalez-Eiras, M. and C. Sanz (2018). Women's representation in politics: The importance of the electoral system. *Unpublished Manuscript*.

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.

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. *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. 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.

Maskin, E. and A. Sen (2016). How majority rule might have stopped Donald Trump. *https://mobile.nytimes.com/2016/05/01/opinion/sunday/how-majority-rule-might-have-stopped-donald-trump.html?referer=*.

McKinney, C. J. (2015). Does the largest political party form the government? *https://fullfact.org/law/does-largest-political-party-form-government/*.

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. *American Economic Review 105*(2), 504–535.

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. (2017). The effect of electoral systems on voter turnout: Evidence from a natural experiment. *Political Science Research and Methods 5* (4), 689–710.

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.

Solé-Ollé, A. and P. Sorribas-Navarro (2017, forthcoming). Trust no more? on the lasting effects of corruption scandals. *European Journal of Political Economy*.

Sweeting, D. (2009). The institutions of strong local political leadership in Spain. *Environment and Planning C: Government and Policy 27* (4), 698–712.

The Guardian (2018). The Five Star Movement has earned the right to govern Italy. *https://www.theguardian.com/commentisfree/2018/mar/07/five-star-movement-govern-italy-corruption-economy-berlusconi*.

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.

Wells, P. (2015). Paul Wells: What if Harper wins a minority? *Maclean's, https://www.macleans.ca/politics/ottawa/paul-wells-what-if-stephen-harper-wins-a-minority/* .

Figure 1a: Effect of Having the Most Seats - National Parliaments Data

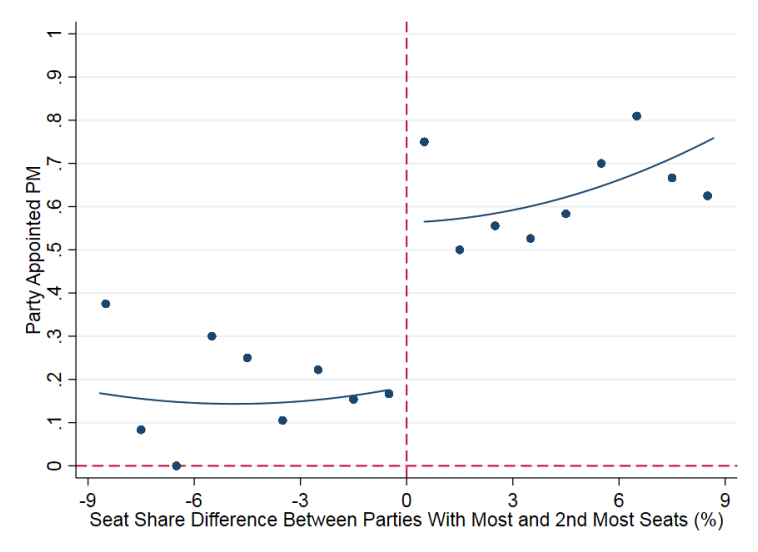
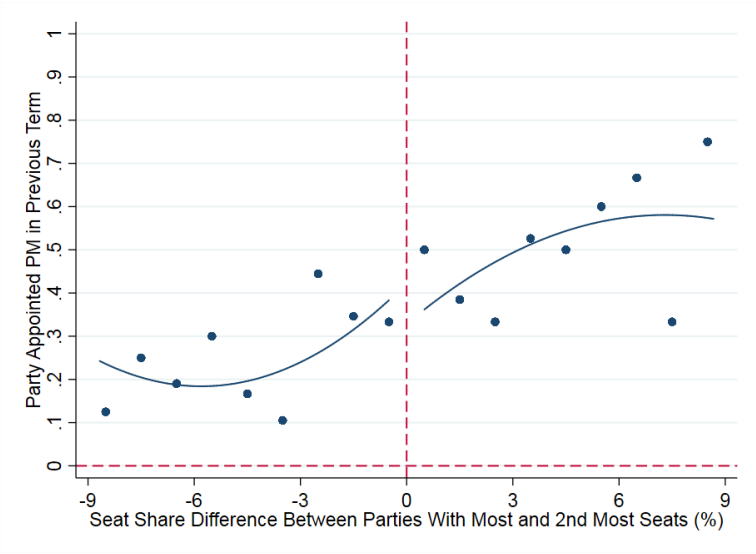


Figure 1b: Placebo Test: "Effect" of Most Seats on Lagged Outcome



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.

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

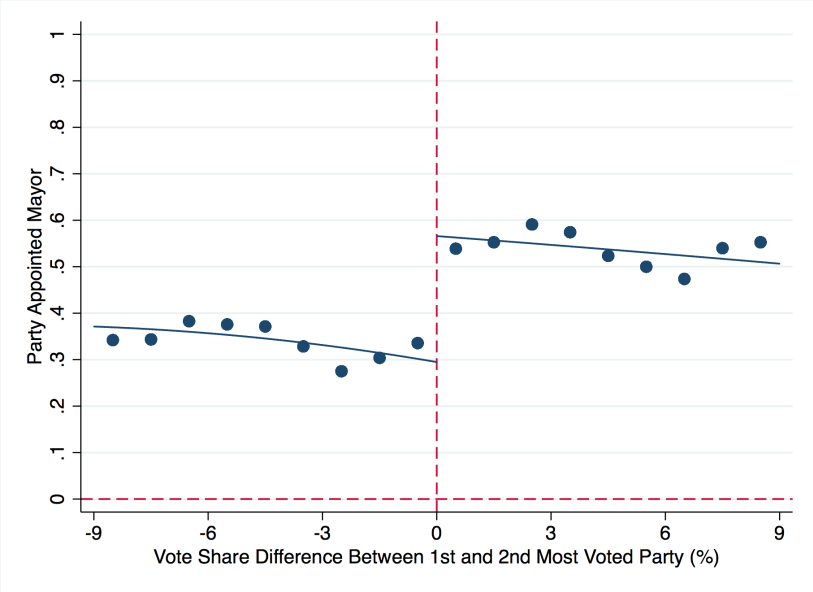
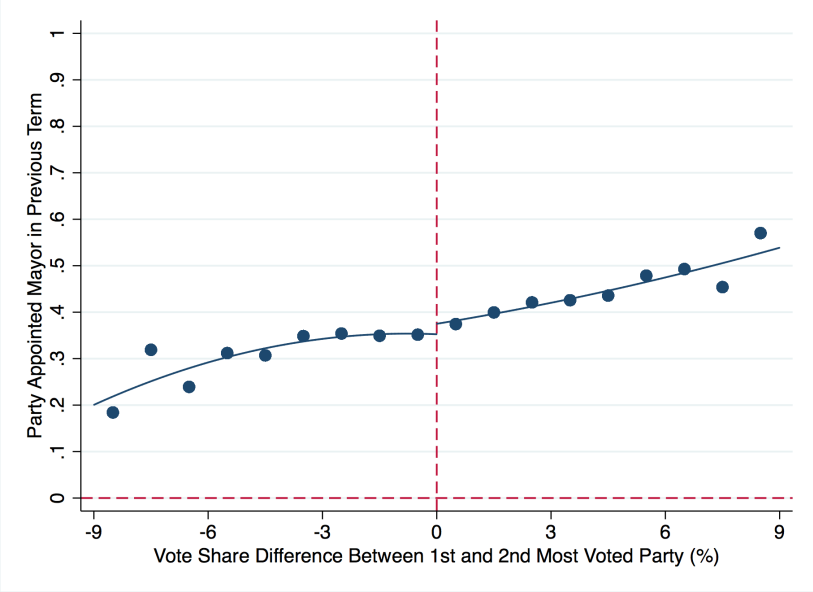


Figure 2b: Placebo Test: "Effect" of Most Voted on Lagged Outcome



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 3a: Effect of Being Second (Instead of Third) Most Voted

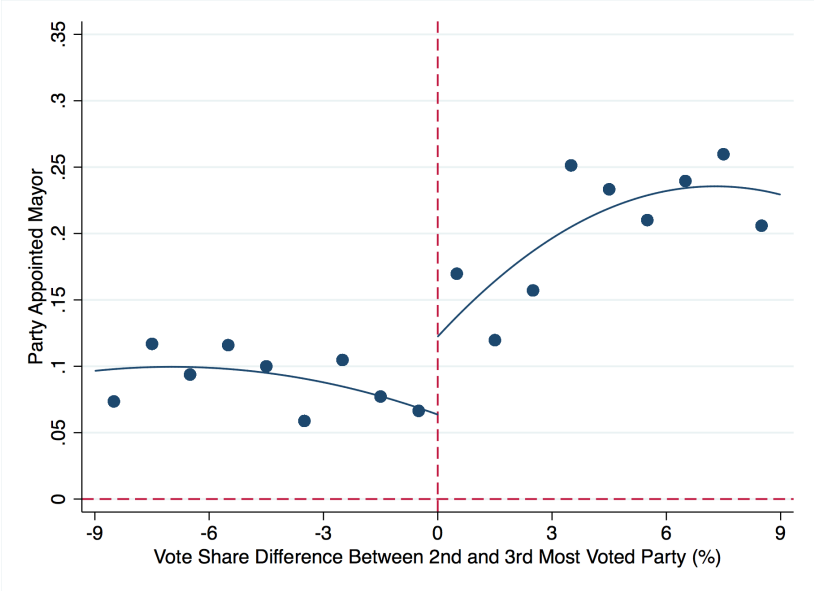
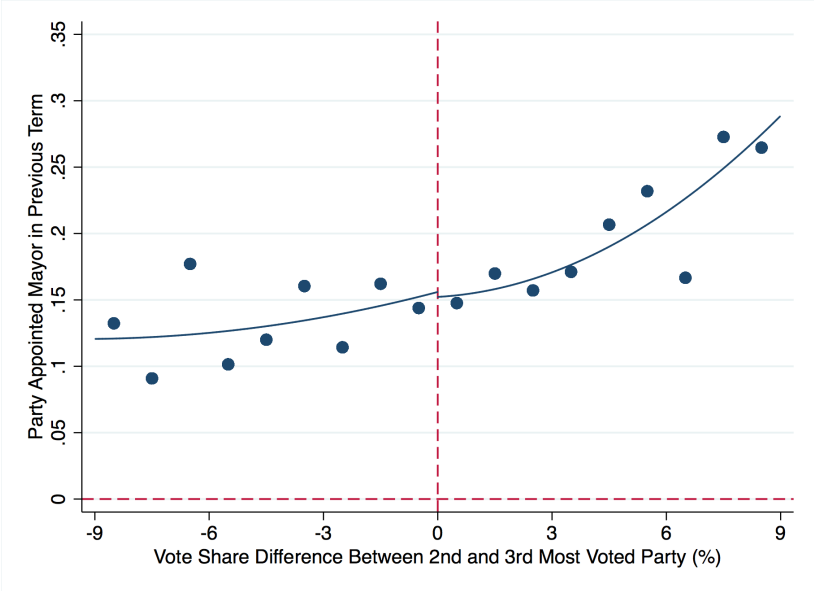
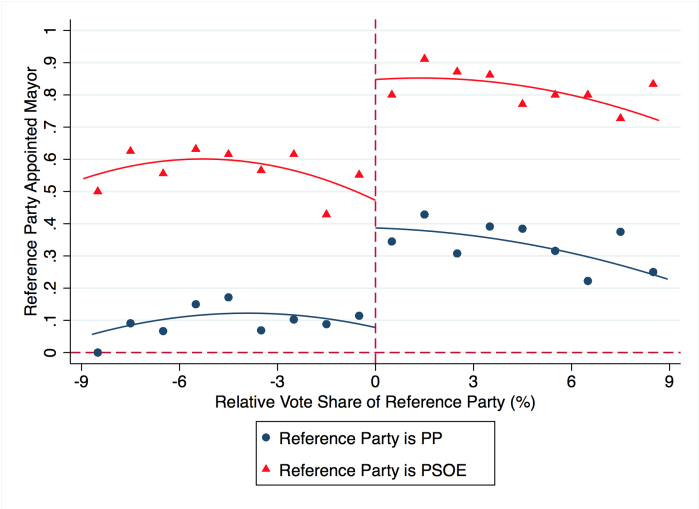


Figure 3b: Placebo Test: "Effect" of Second Most Voted on Lagged Outcome



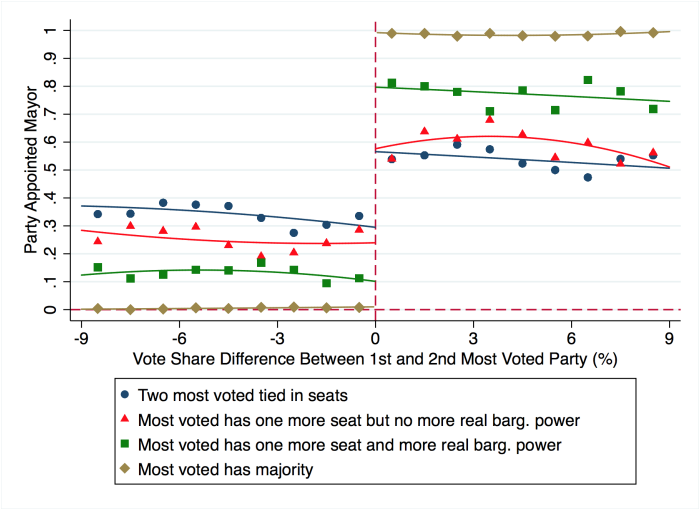
The unit of observation is a party-municipality-year. Sample is restricted to the second and third most voted parties in elections in which they tied in seats and the most voted party did not obtain a majority of seats. The running variable (horizontal axis) is the difference in vote shares between the 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 4: 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 5: 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 6a: Event Study for Effect of Mayoral Appointment on Vote Shares, by Party Rank

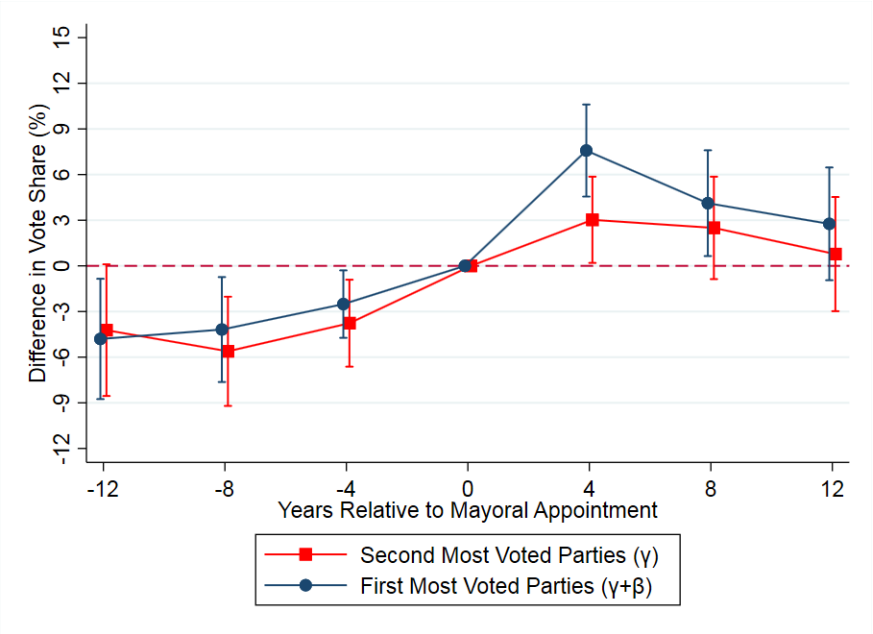
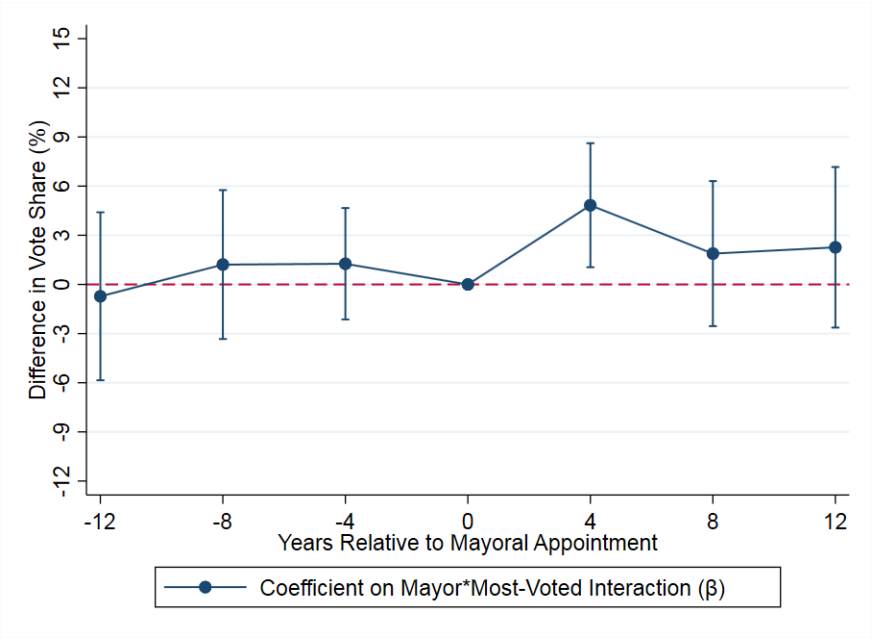


Figure 6b: Event Study for Effect of Mayoral Appointment on Vote Shares, Triple-Differences



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 γ 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 β from equation (2) with different time horizons (see text for further details).

Table 1: National Parliaments Data: Effect of Having Most Seats on Appointing Prime Minister

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

Table 2: Appointment of PMs from the Party with the Most Seats is Associated with Shorter Delays in Government Formation and Longer Governments

	(1)	(2)	(3)	(4)
Panel A: Dep. Var. is log(Days between Election and Government Formation)				
Party with Most Seats Appointed PM	-0.348**	-0.356**	-0.269*	-0.277*
	(0.173)	(0.150)	(0.160)	(0.165)
N	308	308	308	308
Panel B: Dep. Var. is log(Length of Government in Days)				
Party with Most Seats Appointed PM	0.177***	0.210***	0.183**	0.173**
	(0.0674)	(0.0683)	(0.0708)	(0.0722)
N	301	301	301	301
Country FE:		Y	Y	Y
"Legislature Class" FE:			Y	Y
Party ideological family FE:				Y

Robust standard errors in parentheses. The unit of observation is a country-year. Each figure in columns reports a separate regression of the specified dependent variable against a a dummy indicating whether the party with the most seats appointed the prime minister. Column (2) adds a set of country fixed effects. Column (3) adds a set of five dummies indicating each the classes of seat distributions in a legislature described by Laver and Benoit 2015, which capture the relative ability of parties to form coalitions (e.g., a dummy for whether one party has a majority and another for whether any two of the three most represented parties can form a coalition). Column (4) adds a set of 12 dummies indicating the ideological family of the most represented party (e.g., the party being liberal, communist, green). See text for further details.

Table 3: Effect of Being First (Instead of Second) Most Voted

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
Panel A: Main Outcome					
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
Panel B: Lagged Outcome (Placebo Test)					
Party Appointed	0.358	0.011	0.023	-0.015	0.014
Mayor, $t-1$		(0.046)	(0.040)	(0.034)	(0.043)
N		2714	876	5796	5796
Panel C: Covariate Balance					
Party is PP	0.310	0.006	0.009	-0.027	-0.036
		(0.041)	(0.039)	(0.034)	(0.042)
N		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)
N		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 4: Effect of Being Second (Instead of Third) Most Voted

Dependent Variable	3rd-pl. Mean	(1)	(2)	(3)	(4)
Panel A: Main Outcome					
Party Appointed	0.067	0.092**	0.103***	0.059**	0.073**
Mayor		(0.043)	(0.028)	(0.028)	(0.036)
N		888	542	3132	3132
Panel B: Lagged Outcome (Placebo Test)					
Party Appointed	0.149	0.004	0.004	-0.024	0.006
Mayor, $t-1$		(0.037)	(0.034)	(0.034)	(0.045)
N		1844	542	3132	3132
Panel C: Covariate Balance					
Party is PP	0.312	-0.092	-0.044	-0.025	-0.033
		(0.072)	(0.043)	(0.043)	(0.056)
N		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)
N		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 5: 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$_t$*Most Voted$_t$ (β)	4.834**	4.232**	3.382*	5.116*	3.504	-1.263
	(1.931)	(1.881)	(1.814)	(2.909)	(2.512)	(1.735)
Mayor$_t$ (γ)	2.868**	3.091**	2.937**	1.175	4.998**	3.822***
	(1.449)	(1.469)	(1.404)	(2.059)	(2.046)	(1.441)
Most Voted$_t$ (δ)	-1.693	-1.532	-1.399	-0.691	-2.017	-1.184
	(1.393)	(1.355)	(1.304)	(2.035)	(1.898)	(1.121)
Constant	-1.160	-0.681	6.630	0.753	-3.372***	-0.877
	(0.785)	(3.025)	(5.646)	(1.055)	(1.163)	(0.783)
Province FE		Y	Y			
Party FE			Y			
Only elections w. vote share of 3rd > median				Y		
Only elections w. vote share of 3rd < median					Y	
N	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 6: Effect of Being Most Voted on Appointing the Mayor, by Strength of Third-Placed Party

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
Panel A: Third most voted party vote share above median					
Party Appointed	0.290	0.306***	0.290***	0.364***	0.334***
Mayor		(0.078)	(0.056)	(0.048)	(0.062)
N		1014	468	2756	2756
Panel B: Third most voted party vote share below median					
Party Appointed	0.430	0.0402	0.103	0.214***	0.118*
Mayor		(0.088)	(0.066)	(0.059)	(0.070)
N		1014	408	3040	3040
p-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 7: Municipalities Adopting the Norm Have Fewer Instances of Corruption

	Dependent Variable is Corruption Indicator				
	(1)	(2)	(3)	(4)	(5)
$Norm_i$	-0.0253**	-0.0264**	-0.0259**	-0.0253**	-0.0209**
	(0.0117)	(0.0109)	(0.0115)	(0.0118)	(0.0103)
N	2390	2390	2390	2390	2390
Dep. Variable Mean for $norm_i = 0$ cases	0.0693	0.0693	0.0693	0.0693	0.0693
Province FE		Y			
Year FE		Y			
Province × Year FE			Y	Y	Y
Mayor's Party FE				Y	Y
Additional Controls:					Y

Standard errors clustered at the municipality level in parentheses. The unit of observation is a municipality-year. Each column shows the results of a regression of an indicator of whether a corruption case was uncovered in municipality i during the electoral term starting at t against $norm_i$, the share of times that, after an election where the two most voted parties were only 1 p.p. vote share away and tied in seats, the most voted party appointed the mayor. Column (2) adds province and year fixe effects. Column (3) adds a set of dummies for each province-year interaction. Column (4) adds a set of dummies indicating the party of the appointed mayor. Column (5) adds the additional controls: the logarithm of the municipality's population, voter turnout (as a share of registered voters), and a set of four dummies, indicating whether the observation fits in one of the four categories described in Figure 5 (e.g., whether one party has a majority or the top two parties are tied in seats).

Online Appendices - Not for Publication

Appendix A: Effects on the Appointment of Deputy Mayors

To shed light on whether the mayoral appointments are just "symbolic" (as discussed in Subsection 4.2), this appendix reports results on the allocation of deputy mayors.

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.[1]

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 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 A19 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 A7 provides the corresponding estimates.[2] 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.

Appendix B: Further Information on the Theoretical Framework

Proof of Proposition 1. Voters' strategies are a best response since (as discussed previously) a voter that observes a private signal of value i (but before observing election results) expects i to be the most likely state and prefers party of type i to appoint the mayor. This remains true even after the voter conditions its decision on the possibility it may be pivotal (e.g., one party will be one vote away from a majority). 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

[1] 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).

[2] 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 A7 are smaller than in Table 3.

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).

Proof of Proposition 2. Voters' strategies are a best response since (as discussed previously) a voter that observes a private signal of value i (but before observing election results) expects i to be the most likely state and prefers party of type i to appoint the mayor. This remains true even after the voter conditions its decision on the possibility it may be pivotal (e.g., one party will be one vote away from a majority or being most voted). 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.

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 $\boldsymbol{p_t} = [p_t^A, p_t^B, p_t^C]$ denotes the probabilities of state $s_t \in \{A, B, C\}$ occurring. In particular, assume that $G(\boldsymbol{p}_t)$ is such that its possible realizations are i) $\boldsymbol{p}_t = [0.45, 0.35, 0.20]$, ii) $\boldsymbol{p}_t = [0.35, 0.45, 0.20]$, and iii) $\boldsymbol{p}_t = [0.20, 0.20, 0.60]$. 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 σ 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 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 $\boldsymbol{p}$ 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 $G(\boldsymbol{p}_t)$ 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 $G(\boldsymbol{p}_t)$ were instead $[0.40 + \epsilon, 0.40 - \epsilon, 0.20]$ and $[0.40 - \epsilon, 0.40 + \epsilon, 0.20]$, with $\epsilon \to 0$, we have a case where parties A and B almost tie and the updating works similarly.

Appendix C: 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.[3]

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

[3] A proposer keeps y and offers one randomly drawn other party $1 - y$. For the other party to accept, its payof 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 $y = \frac{2}{3}$.

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).

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 D: The Role of the 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 4. 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 4.3 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 (Subsection 6.2) or the size of municipality or council (Figure A9).

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 Subsection 4.1, 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 Subsection 4.1, 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, we outline below 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.

Lastly, it should be noted that there is little available data on whether the status quo rule was applied or not (e.g., whether a mayor was appointed through a majority of votes or not). However, we believe there would be little meaningful information in this data. For example, suppose that, in a

three party legislature, each party votes for their own leader, and the mayor is appointed by the status quo rule. For both the second and third most voted parties, unilaterally deviating from this strategy and voting for the most voted mayors does not affect the final outcome (who is the mayor) but it does change whether the status quo rule is applied or not. Given there is indifference in equilibrium, there is little information to be learned about parties incentives and beliefs from this choice. Moreover, the voting game outlined below illustrates that this indifference argument applies in all the equilibria where the most voted party appoints the mayor.

The role of the status quo 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 Subsection 4.1.

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 4. 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, ϕ, a), (a, a, ϕ), (a, ϕ, ϕ), (a, b, b), (a, c, c), (b, b, b), (c, c, c), (ϕ, a, a), (ϕ, ϕ, a), (ϕ, a, ϕ), (ϕ, ϕ, ϕ), (ϕ, b, b), and (ϕ, 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 ϕ) for party A.[4] Additionally, a and ϕ are dominated by b (for party B) and c (for party C).[5] Hence, there are only four Nash equilibria where a party is not playing a weakly dominated strategy: (a, b, b), (a, c, c), (ϕ, b, b), and (ϕ, 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.

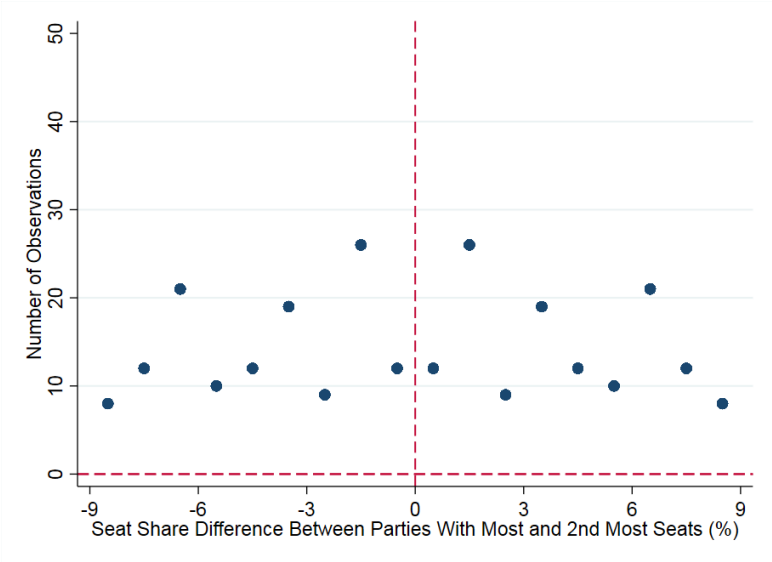
To illustrate the irrelevance of the status quo status further, one could reanalyze the game but now making party B the status quo. Following a similar argument, it can be shown there is no equilibrium where A appoints the mayor and players do not choose a weakly dominated strategy. Hence, A's ability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

[4] If the other two parties are not casting the same vote, A and ϕ can guarantee the best outcome for A. If only one other party is playing a or ϕ, 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.

[5] 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 ϕ. A similar argument applies to party C.

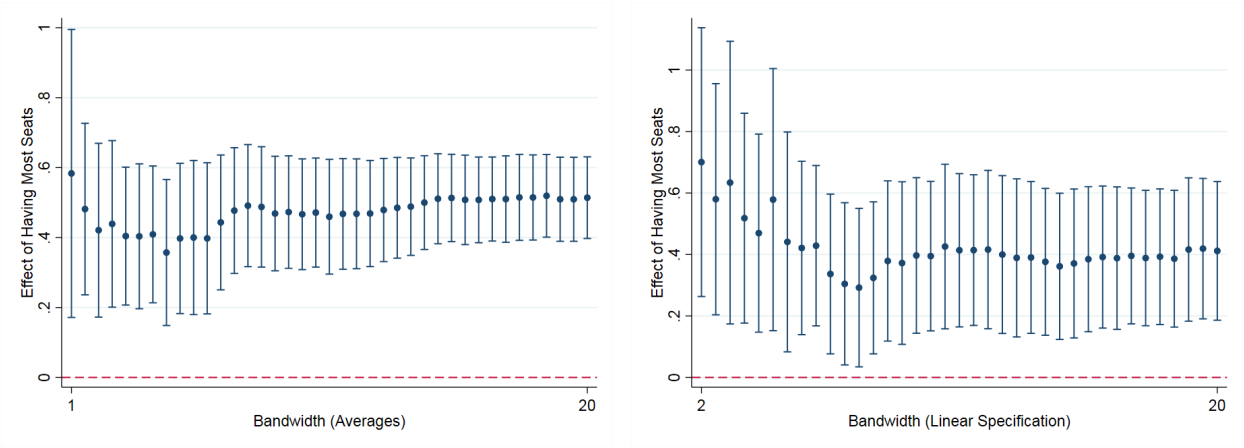
Appendix Figures and Tables

Figure A1: Histogram - National Parliaments Data



The unit of observation is a country-year-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 A2: Robustness to Bandwidth Choice - National Parliaments Data



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

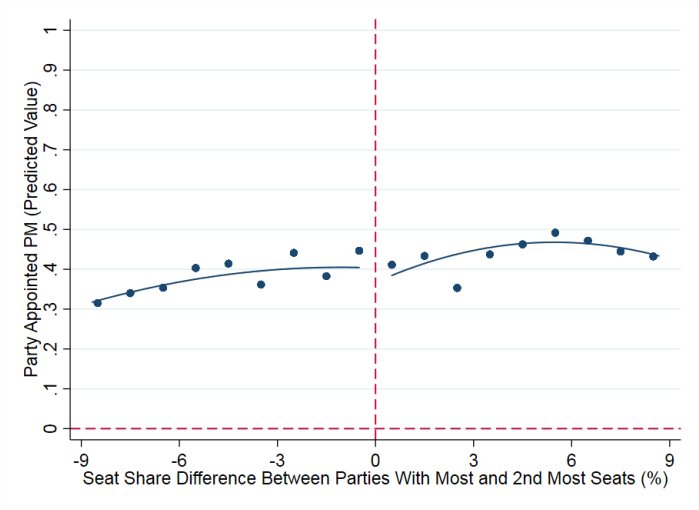
(b) Effect of 1st vs. 2nd (Specification: Linear)

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

Figure A3: Effect of Having Most Seats on Being in Ruling Coalition - National Parliaments Data

The unit of observation is a country-year-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.

Figure A4: Covariate Balance: National Parliaments Data

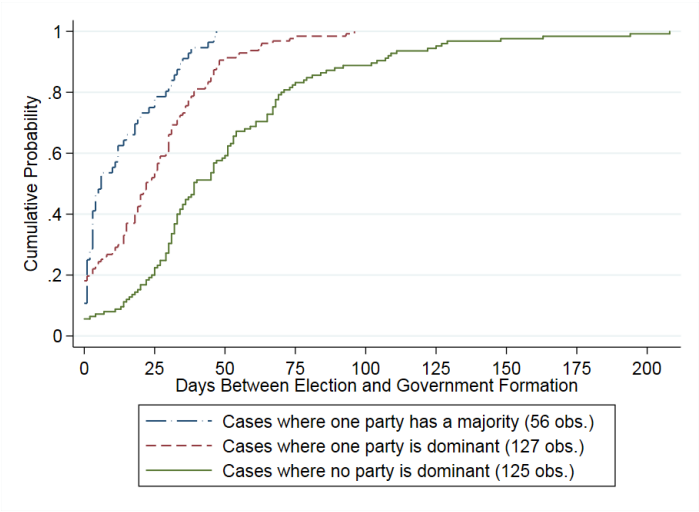


The unit of observation is a country-year-party. Sample is restricted to the two parties with the most seats in the parliament. The running variable (horizontal axis) is the difference in *seat* shares between the two most voted parties: positive for the party with the most seats and negative for the 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 A5: Effect of Having Most Seats in Non-"Dominant" Cases - National Parliaments Data

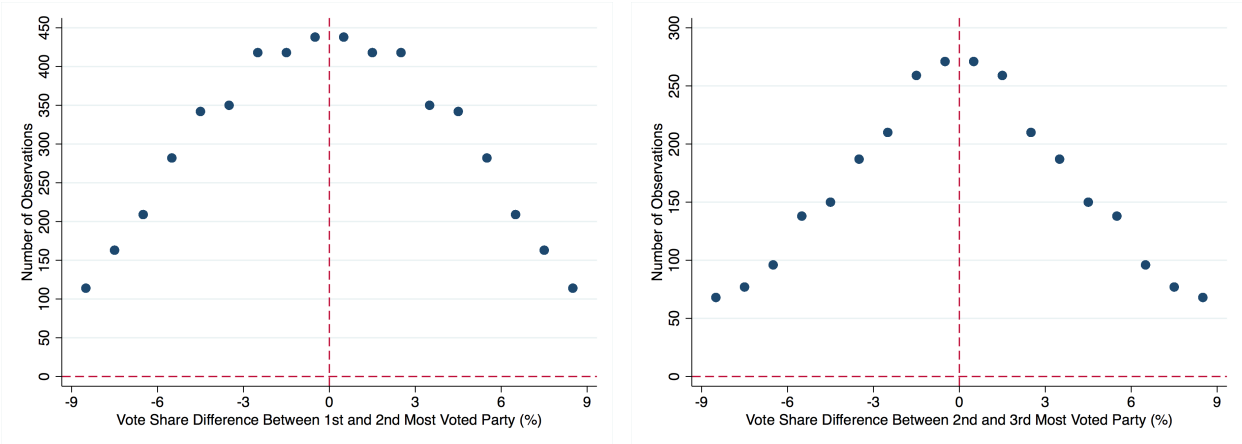
The unit of observation is a country-year-party. Sample is restricted to the two parties with the most seats in parliaments with non-"dominant" seat compositions (it excludes 51% of elections where the first and third placed parties can form a majority, while the second and third placed cannot). The 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 A6: CDF of Bargaining Delays in Government Formation - National Parliaments Data



The unit of observation is a country-year. The figure plots the cumulative distribution function (CDF) of the time between election and government formation for the cases where: i) one party in the legislature has a majority of votes, ii) the party with most seats is "dominant" (i.e., can form a majority with the party with the third most seats, while the second most voted cannot do the same), iii) no party is dominant.

Figure A7: Histograms: Distribution of Running Variable

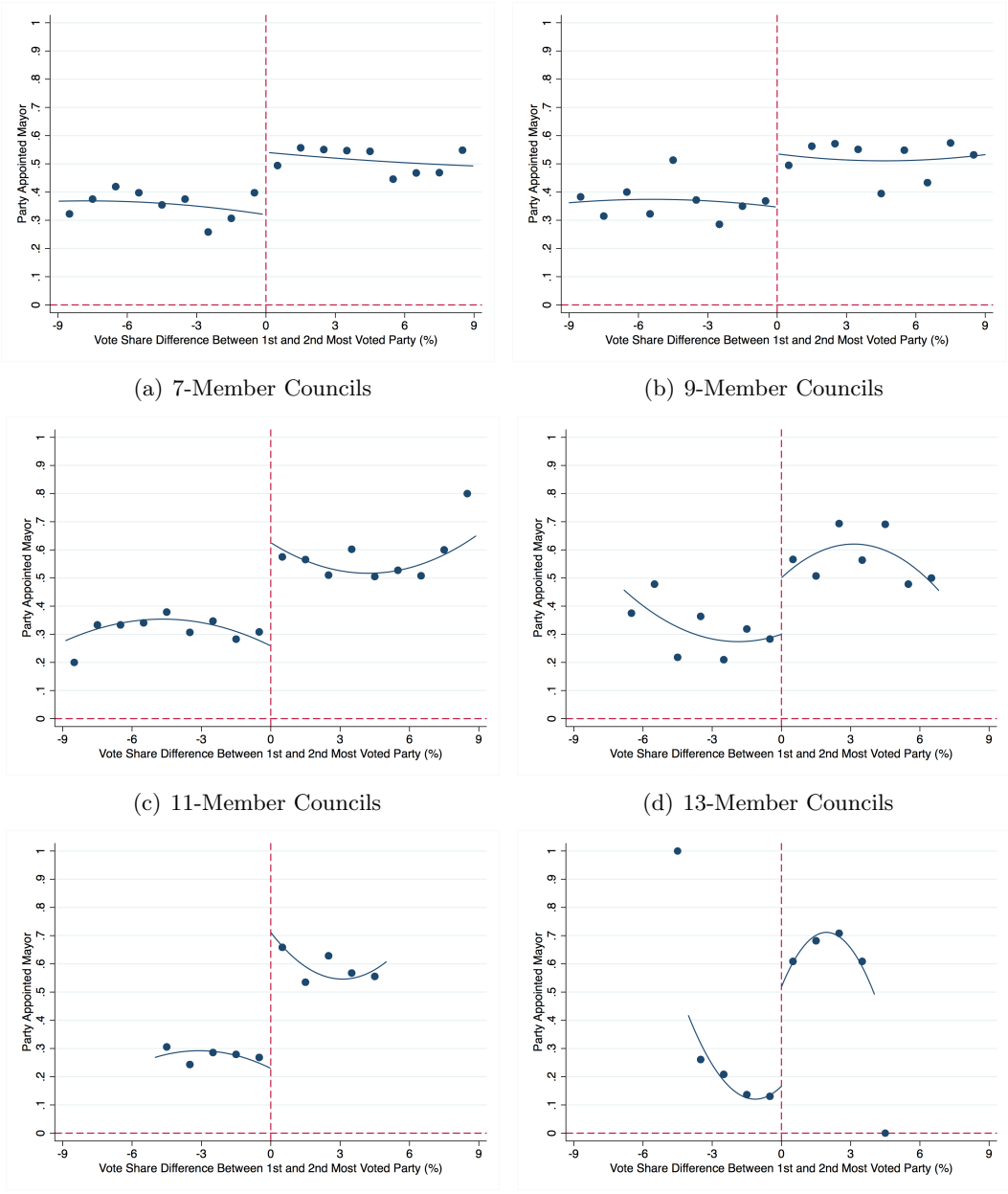


(a) 1st-vs.-2nd Histogram

(b) 2nd-vs.-3rd Histogram

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 A8: Effect of Being Most Voted, Heterogeneity by Council Size



(a) 7-Member Councils

(b) 9-Member Councils

(c) 11-Member Councils

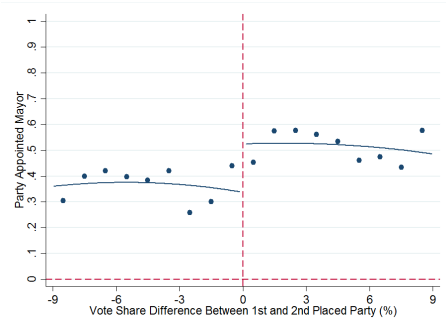
(d) 13-Member Councils

(e) 17-Member Councils

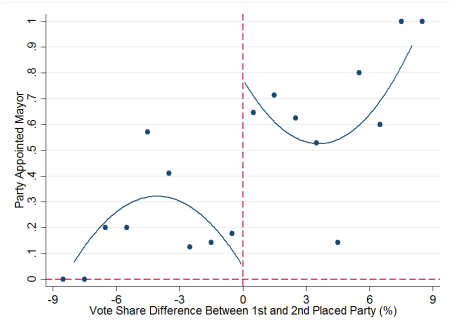
(f) 21-Member Councils

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).

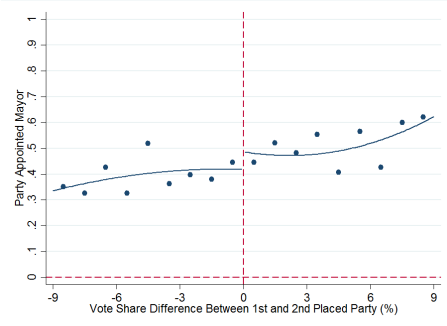
Figure A9: Effect of First-Place, by Council Type



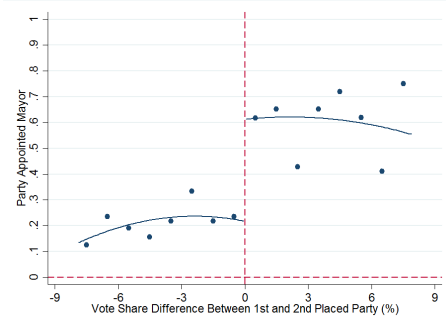
(a) 3-3-1 Councils



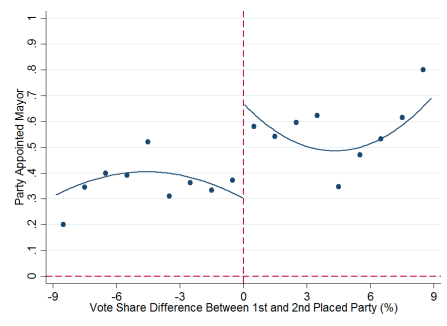
(b) 3-3-3 Councils



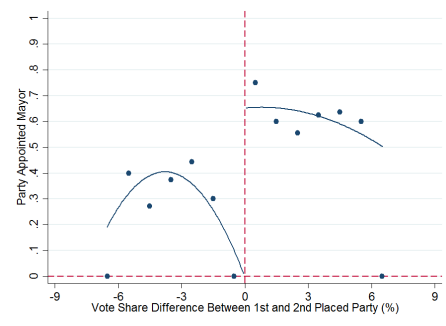
(c) 4-4-1 Councils



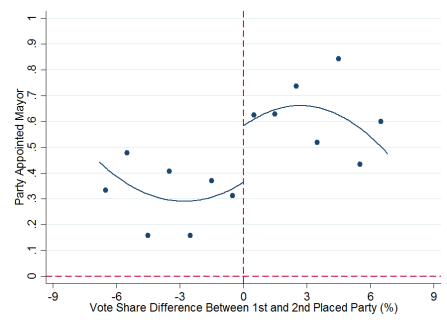
(d) 4-4-3 Councils



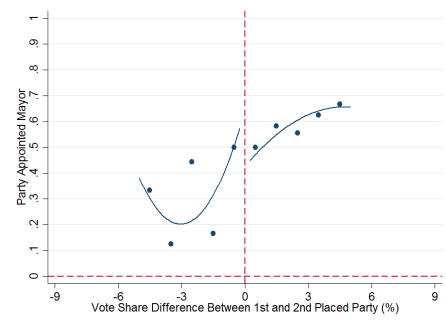
(e) 5-5-1 Councils



(f) 5-5-3 Councils



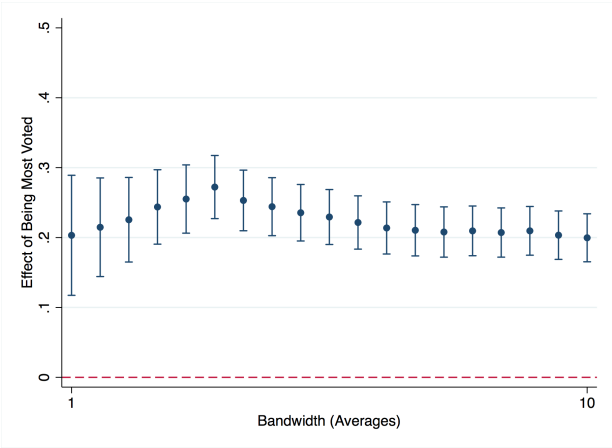
(g) 6-6-1 Councils



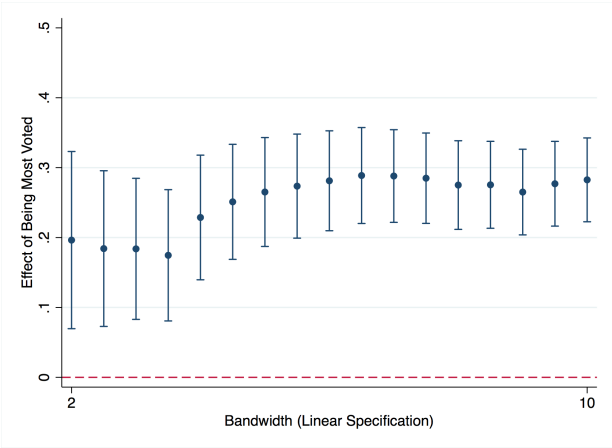
(h) 8-8-1 Councils

Notes are same to those on Figure A8, 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).

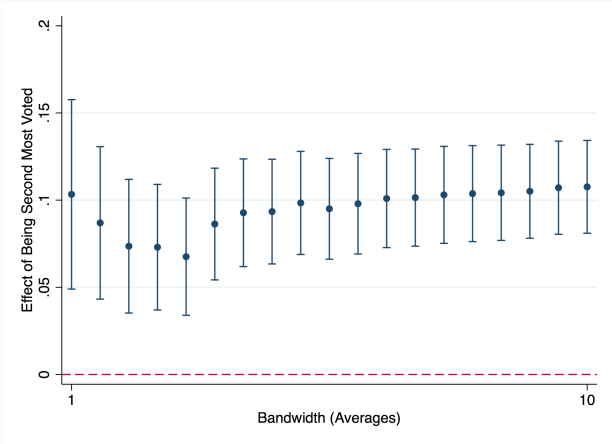
Figure A10: Robustness to Bandwidth Choice.



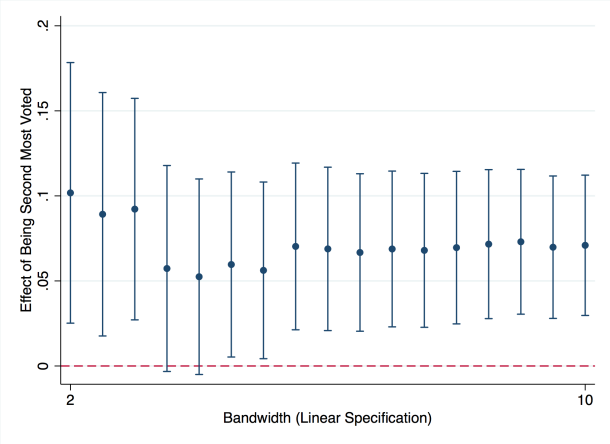
(a) Effect of 1st vs. 2nd (Specification: Means)



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



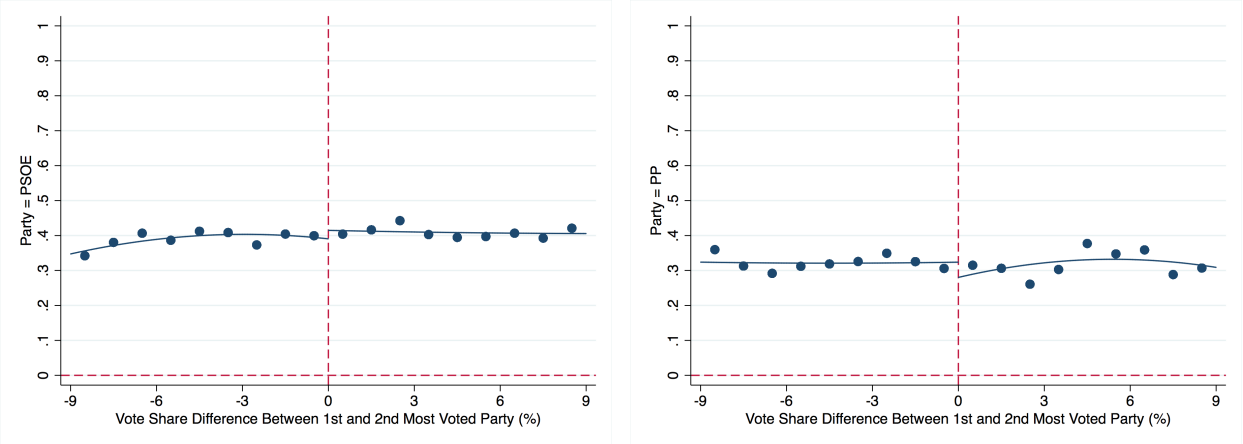
(c) Effect of 2nd vs. 3rd (Specification: Means)



(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 A11: Covariate Balance (1st vs. 2nd): Placebo "Effect" on Party Identity

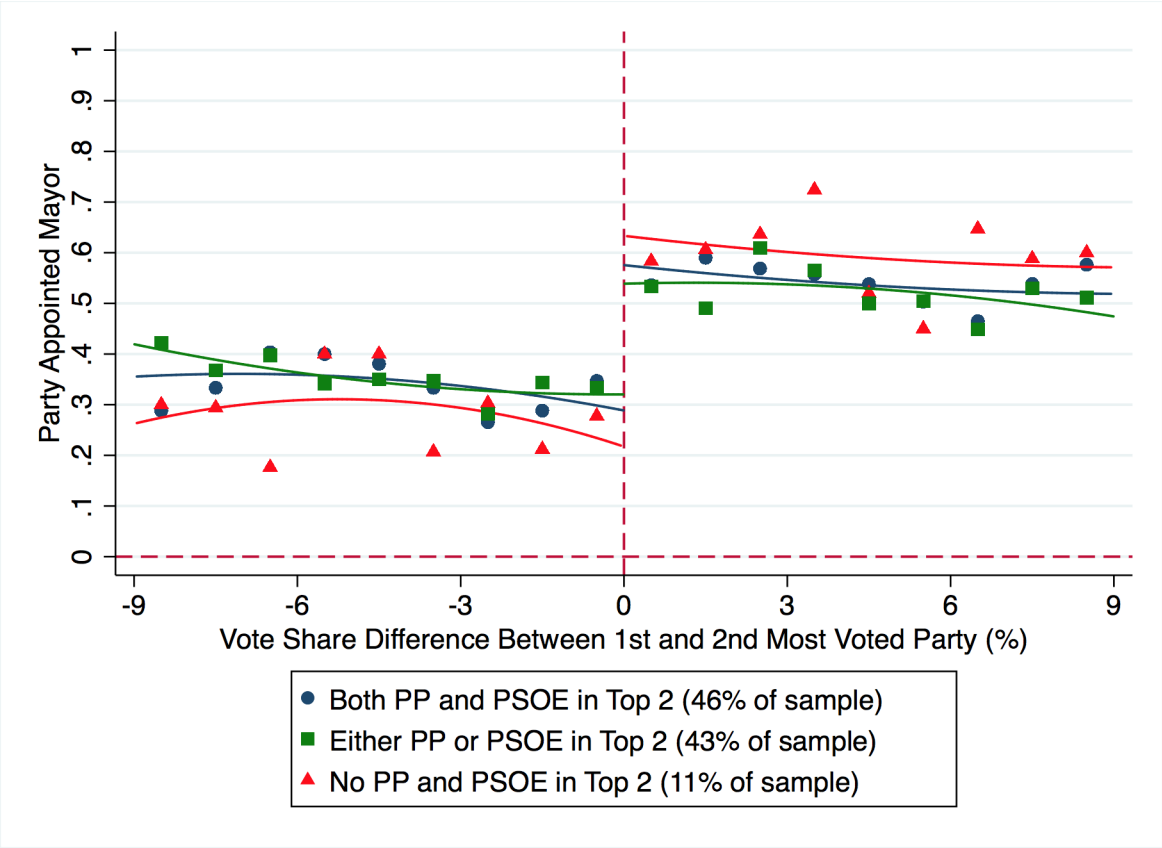


(a) Indicator for party being the PSOE

(b) Indicator for party being the PP

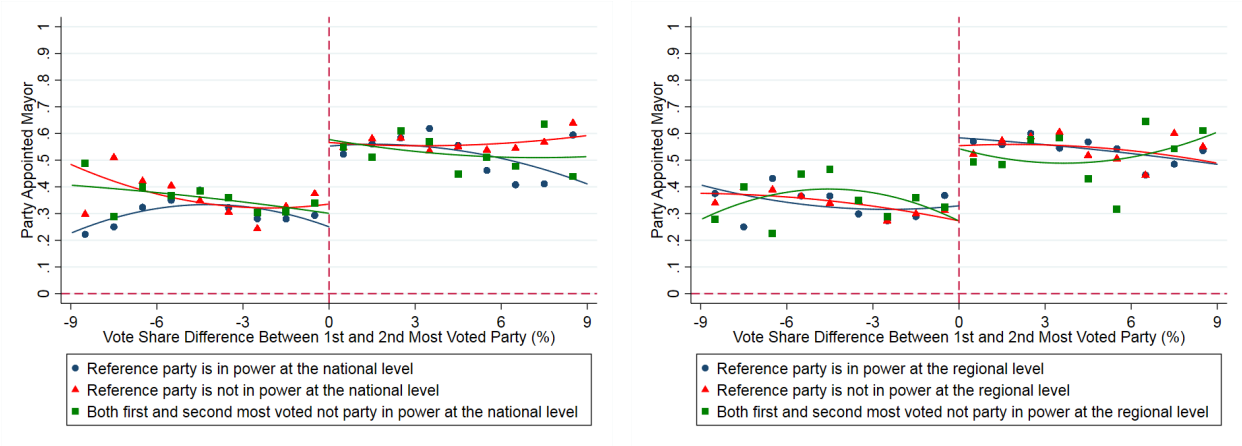
The unit of observation is a party-municipality-year. Sample is restricted to the two most voted parties in elections in which they tied in seats. The running variable (horizontal axis) is the difference in vote shares between the two most voted parties: positive for the most voted party and negative 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 A12: 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.

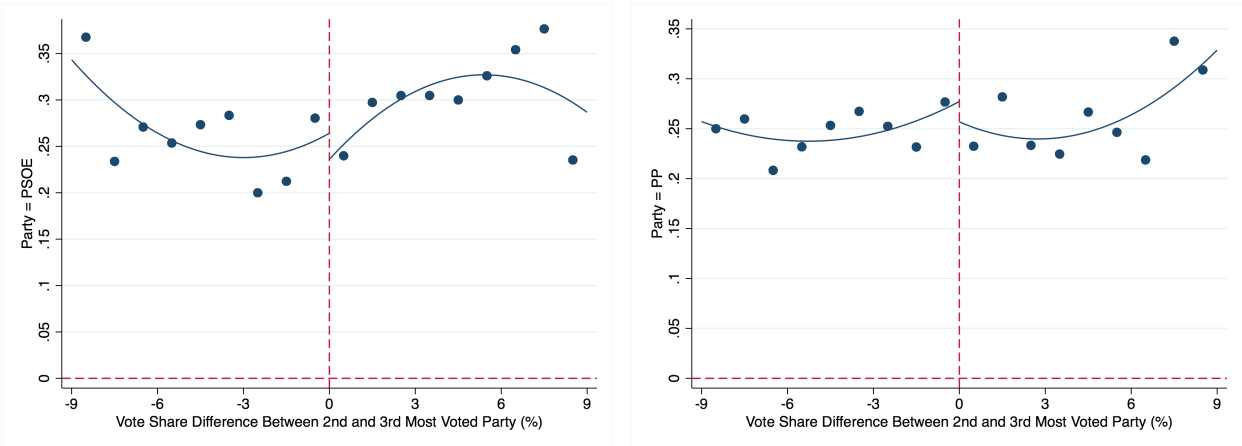
Figure A13: Is the Most Voted Effect Different for Powerful Parties?



(a) National (b) Regional

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 (B), the three separate plots are for the cases where the reference was in power at the national (regional) level, was not in power at that level, and in which both parties were not in power at that level.

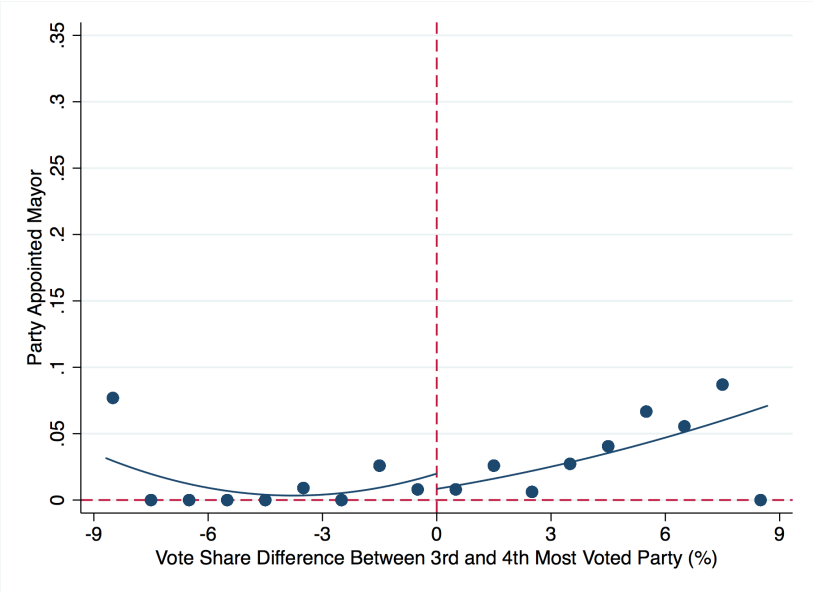
Figure A14: Covariate Balance (2nd vs. 3rd): Placebo "Effect" on Party Identity



(a) Observation belongs to the PSOE (b) Observation belongs to the PP

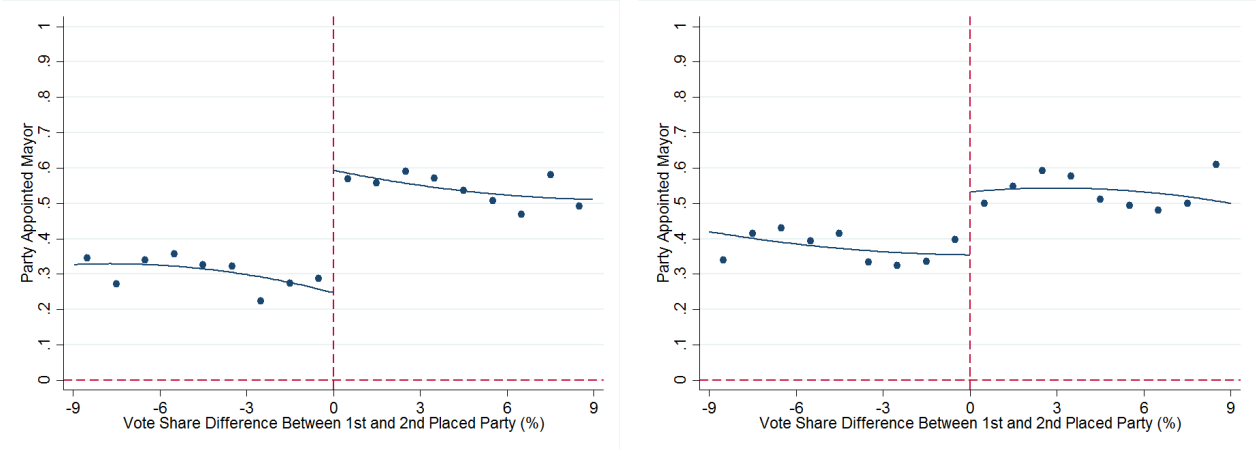
The unit of observation is a party-municipality-year. Sample is restricted to the second and third most voted parties in elections in which they tied in seats and the most voted party did not obtain a majority of seats. The running variable (horizontal axis) is the difference in vote shares between the 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 A15: 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 A16: Effect Heterogeneity by Third-Placed Party Vote Share

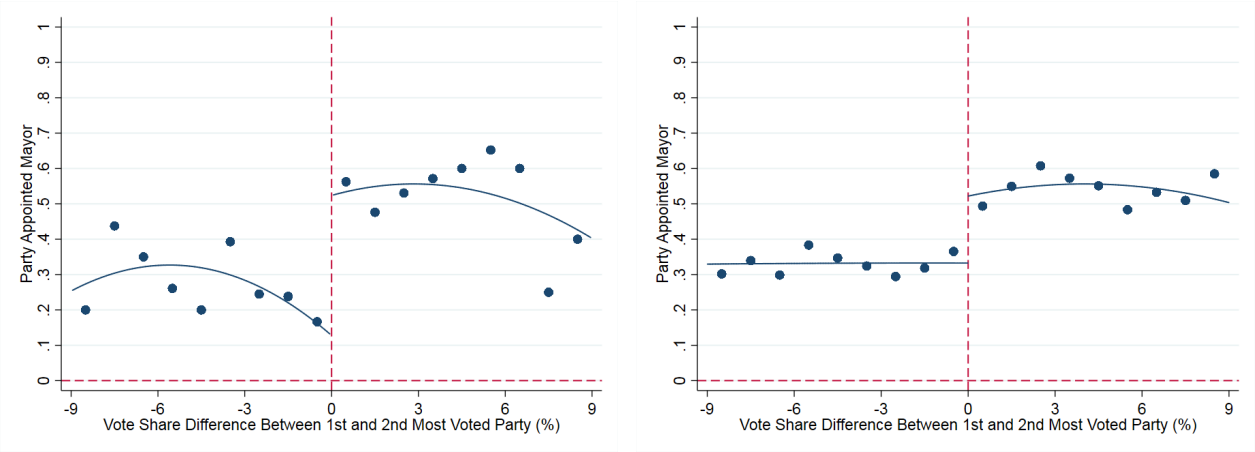


(a) Third placed party vote share above median

(b) Third placed party vote share below median

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 A17: Effect Heterogeneity by Whether Third-Placed Party was (or will be) the Most Voted

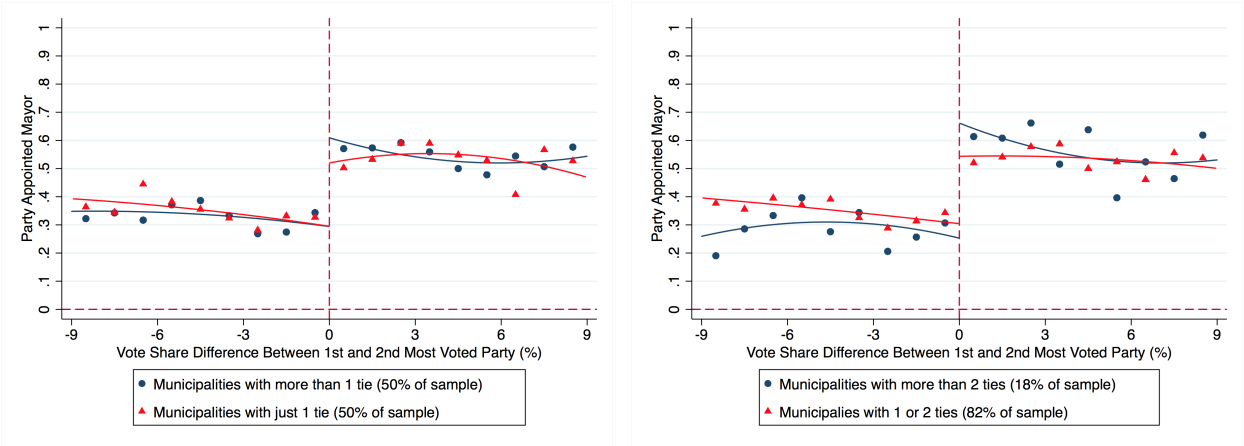


(a) Third placed party vote is the most-voted in at least one of he next (or last) three elections

(b) Third placed party vote is not the most-voted in any of the next (or last) three elections

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-placed party has ever (never) been the most voted party at any of the three previous or subsequent elections.

Figure A18: Effect Heterogeneity by Frequency of Ties in Municipality

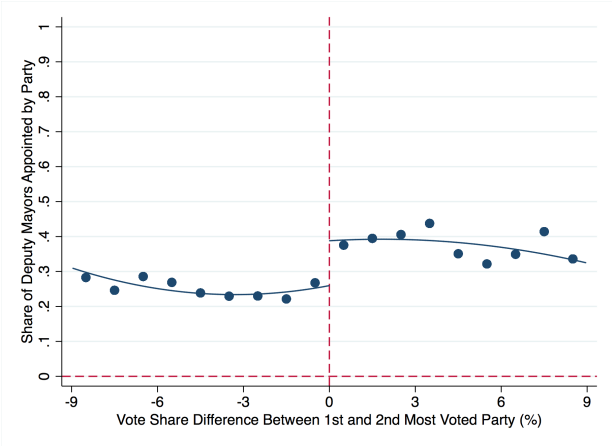


(a) Effect of Most Voted, by Frequency of Ties

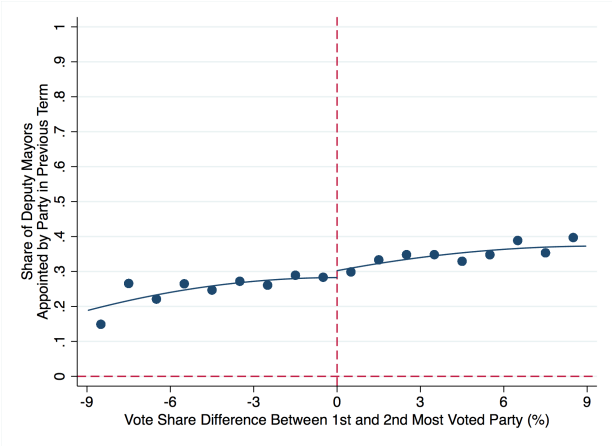
(b) Effect of Most Voted, by Frequency of Ties

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.

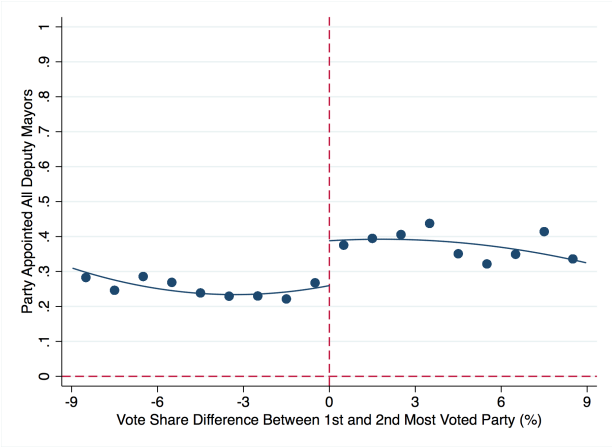
Figure A19: Effect of Being Most Voted on Deputy Mayors' Allocation



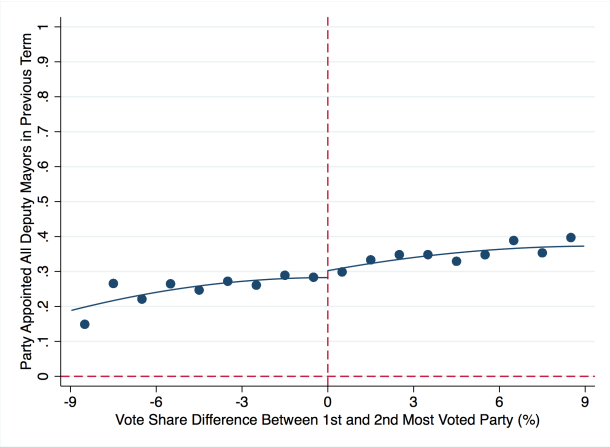
(a) Effect of Most Voted on Share of Deputy Mayors



(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.

Table A1: Distribution of Council Sizes

		Number of Municipality-Elections		
Population	Number of Seats	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)
Panel A: Main Outcome (mayor serves at least 3/4 of term)					
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
Panel B: Outcome is appointing mayor for entire term					
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
Panel C: Outcome is appointing mayor for longer than other parties					
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
Panel D: Outcome is appointing initial mayor					
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
Panel E: Main outcome, sample restricted to cases where "two parties out of top-3 needed for majority"					
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 First (Instead of Second) Most Voted: Including 1979 and 2015 Elections

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
Panel A: Main Outcome (mayor serves at least 3/4 of term), with 1979 data					
Party Appointed	0.343	0.209***	0.204***	0.290***	0.242***
	(0.0621)	(0.0417)	(0.0347)	(0.0437)	
N		1882	980	6474	6474
Panel B: Outcome is appointing mayor for entire term, with 1979 data					
Party Appointed	0.330	0.205***	0.218***	0.300***	0.254***
	(0.0551)	(0.0405)	(0.0339)	(0.0427)	
N		2206	980	6474	6474
Panel C: Outcome is appointing mayor for longer than other parties, with 1979 data					
Party Appointed	0.366	0.223***	0.220***	0.304***	0.267***
	(0.0610)	(0.0427)	(0.0351)	(0.0447)	
N		2050	980	6474	6474
Panel D: Outcome is appointing initial mayor, with 1979 data					
Party Appointed	0.360	0.244***	0.245***	0.335***	0.288***
	(0.0565)	(0.0424)	(0.0352)	(0.0445)	
N		2282	980	6474	6474
Panel E: Outcome is appointing initial mayor, with 1979 and 2015 data					
Party Appointed	0.374	0.222***	0.233***	0.310***	0.267***
	(0.0543)	(0.0409)	(0.0341)	(0.0431)	
N		2464	1074	7004	7004
Panel F: Main outcome, sample restricted to cases where "two parties out of top-3 needed for majority", with 1979 data					
Party Appointed	0.357	0.193***	0.205***	0.295***	0.244***
	(0.0647)	(0.0443)	(0.0364)	(0.0459)	
N		1790	888	6110	6110
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 A4: Effect of Being Most Voted: Cases with a Left-Wing Majority

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
Panel A: Effect for PSOE (conditional IU being third most voted)					
PSOE Appointed	0.543	0.267*	0.248**	0.417***	0.264*
Mayor		(0.153)	(0.118)	(0.109)	(0.145)
N		155	64	423	423
Panel B: Effect for PP (conditional IU being third most voted)					
PP Appointed	0.102	0.242*	0.231**	0.313***	0.246*
Mayor		(0.146)	(0.110)	(0.101)	(0.139)
N		155	64	423	423
p-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 A5: Comparing Magnitude of Effects:
Effect of Being Most Voted, by Legislature Type

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
Panel A: First and second most voted tied in seats					
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
Panel B: Most voted has one more seat than second most voted, but no more "real" bargaining power					
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
Panel C: Most voted has one more seat than second most voted and also more "real" bargaining power					
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
Panel D: Most voted has a majority of seats					
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.

Table A6: Effect of Being Most Voted on Appointing the Mayor, by Probability of Third-Placed Becoming the Most Voted

Dependent Variable	2nd-pl. Mean	(1)	(2)	(3)	(4)
Panel A: Third most voted party vote is the most voted at some election (t-3, t+3)					
Party Appointed	0.170	0.494***	0.396***	0.351***	0.453***
		(0.152)	(0.111)	(0.117)	(0.153)
N		218	96	550	550
Panel B: Third most voted party vote is never the most voted (t-3, t+3)					
Party Appointed	0.400	0.0541	0.128*	0.170**	0.111
		(0.0976)	(0.0738)	(0.0689)	(0.0946)
N		804	312	2298	2298
p-value: test of equal effects		0.0137	0.0411	0.1818	0.0543
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 ever (never) been the most voted party at any of the three previous or subsequent elections. 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 A7: Effect of Being Most Voted on Deputy Mayors' Allocation

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