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 being labeled the most voted. This effect is comparable to that of obtaining an additional seat, and is also present when a right-wing party is the most voted and the second and third most voted parties are allied left-wing parties who can form a combined majority. 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.
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, informal rules, and conventions about 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. We provide evidence that bargaining outcomes are consistent with 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).

Politicians commonly appeal to this prescription as “democratic” or “respecting the will of the people.” For example, Canadian prime minister Stephen Harper stated in an interview 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.” When the interviewer pressed him by asking “So, even as the current government, if you’re just a couple of seats behind, you wouldn’t try to figure out a way to…”, 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).

Although Mr. Harper’s states that a plurality (“the most seats”) legitimizes a government, formal Canadian institutions are majoritarian: a prime minister requires the support of more than half of parliament. Furthermore, Mr. Harper’s motivations are not directly attributable to his self interest, as his party was not expected to win the most seats. Politicians and voters in other contexts make arguments similar to Mr. Harper’s. The next section provides examples from New Zealand, United Kingdom, Italy, and Spain.

This example illustrates a broader issue that Maskin and Sen (2016) refer to as the “serious confusion when a plurality win is marketed as a majority victory” and their argument that “understanding of the critical difference between a plurality and a majority could improve politics around the world.” Maskin and Sen’s (2016) point goes beyond government formation, extending to issues as the legitimacy of the BJP’s hindu-nationalism in India, the effects of the Muslim Brotherhood on Egypt’s democratic institutions, and Donald Trump’s nomination as presidential candidate in the USA.

To study the informal prescription that the “party with most seats should form government,” we first 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 focus on Spanish municipalities, which provide a sharper identification strategy. Each municipality elects a council by (closed list) proportional representation under the D’Hondt apportionment rule in a single-district election. In its first meeting, the council selects, by majority

---

1Proportional 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.
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 number seats, exploiting a 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.

As an example, consider a council with 11 seats where 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. In theory, A and B should have equal bargaining power and be equally likely to form a government with majority support. Thus, the probability A and B appoints the mayor should be the same. Our findings indicate that this probability is 55% for party A and 35% for party B.

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. While 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.

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. Comparing most voted parties to lower ranked ones in general (i.e., including cases where parties’ vote shares differ substantially) is potentially confounded by the different number of seats and other factors that lead one party to outperform others. However, these are held constant in our RDD.

Additionally, by focusing on parties that almost tied in votes, our results indicate that agents respond to 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; Folke, Persson, and Rickne 2016; Meriläinen and Tukiainen 2018; Pons and Tricaud 2018) and is consistent with models where agents make choices over-weighting salient attributes (Bordalo, Gennaioli, and Shleifer 2013). Our results thus speak to how narrow victories are interpreted (or marketed) by some as “the will of the people.” For example, only 51.9% of votes in a British 2016 referendum were for leaving the European Union and only 50.6% of the votes in a 1995 Quebec referendum were against independence from Canada. These results went on to be interpreted (by some) as “what the people has chosen.” After becoming the US president with 50.7% of the vote, George W. Bush justified his agenda by saying “the people made it clear what they wanted” and “I earned capital in the campaign, political capital, and now I intend to spend it” (Stevenson 2004).

Our results are difficult to reconcile with existing theories of legislative bargaining and government

---

1Only party leaders can become mayor. Thus, in our example only three members of the 11-member council can be mayor. Section 4 describes these electoral rules in detail.

2Sections 4 and 7 discuss additional reasons why the status quo rule cannot explain our results. Mayors require continuous support from a majority during the term. In our example, parties B and C could appoint a B mayor anytime they agree to do so, making it less likely the status quo binds. We also provide evidence consistent with voters punishing second most voted parties that appoints mayors. It is unclear why the status quo rule would generate this result.
formation. Their starting point is the number of seats held by voting blocks (political parties). Given legislative procedures, seat distributions fully determines which coalitions can be formed and hence parties’ bargaining power. Such theories thus take parties’ seats as the primitives and ignore the role of the votes that lead to these seat allocations: seat distributions should be sufficient to study bargaining outcomes.\footnote{This applies to both noncooperative and cooperative theories (e.g., core or bargaining sets). Ray and Vohra (2014) survey coalition formation and Laver (1998) surveys government formation in particular. Examples of models of legislative bargaining and government formation are Baron and Ferejohn (1989), Austen-Smith and Banks (1990), Laver and Shepsle (1986), Baron (1991, 1993), Merlo and Wilson (1995), Morelli (1999), and Snyder Jr, Ting, and Ansolabehere (2005).} Seat allocations are also the focus of empirical work. For example, Gamson’s Law is the empirical regularity that cabinet positions are distributed proportionally to coalition parties’ contribution of seats. Structural models of government formation also take seat allocations as the starting point. Indeed, standard datasets used in this literature usually contain information only on parties’ seat allocations (and not their general election votes).\footnote{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, parties act in a manner consistent with the prescription even when it goes against programmatic affinity between parties. 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 most-voted right-wing party frequently appoints the mayor instead.\footnote{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.\footnote{Persson, Roland, and Tabellini (1997) also interpret different equilibria in voting models as different norms.}}

To gauge the magnitude of the effect of being most-voted, we compare it to the effect of obtaining one additional seat, which can also be identified in a RDD. Having a plurality (but not a majority) of seats has an effect only slightly larger than being the most voted (but tied in number of seats). This suggests that the importance of the novel effect we document relative to previously studied determinants of bargaining outcomes.

While these results can appear as prima facie irrational or puzzling, we present a simple model that can rationalize them. 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 to appoint the mayor. However, parties’ representation in a council is set at this point, and bargaining over mayoral appointments can be based on factors that ignore voters’ interests. This creates a conflict between voters and parties and a role for disciplining. The model has multiple equilibria, which can be interpreted as different norms (self-enforcing rules of behavior) that voters can adopt.\footnote{Persson, Roland, and Tabellini (1997) also interpret different equilibria in voting models as different norms.} A norm that matches our results and where voters punish second most voted parties that appoint the mayor, constitutes an equilibrium (that also maximizes voters’ expected welfare, suggesting an instrumental value for the norm).

We provide three pieces of evidence consistent with the model. First, we test whether voters’ behavior is consistent with punishing parties that deviate from the norm. We use a triple-difference
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. Some alternative explanations for our results (e.g., the status quo rule or the effect arising as an agreement among parties in repeated bargaining) do not naturally lead to these. Third, we provide evidence suggesting that the norm we study affects policymaking: Spanish municipalities that appoint most-voted mayors have fewer instances of government corruption and, in European countries, appointing the prime minister from the party with most seats is associated with shorter bargaining delays and longer government duration.

This paper is related to four strands of the literature. First, as previously discussed, our result is difficult to reconcile with existing theories of bargaining and coalition formation and suggests the importance of a relatively unexplored factor in bargaining outcomes.

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 results we document add first-past-the-post considerations to proportional systems. For example, [Lizzieri and Persico (2001)] compare how public good provision differs under different electoral systems. In their model, parties under proportional representation maximize vote shares (which translates proportionally to power), while parties under plurality rule maximize the probability of being most voted (a winner-takes-all contest). At first pass, the Spanish context fits the proportional representation case. However, our results suggest this characterization misses that parties should not simply maximize vote shares but also aim at being the most voted. 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 results we document can generate equivalent de facto premia even without such explicit rules.

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 (previously discussed) evidence that rank-based decision-making affects political outcomes. Fourth, our results speak to the literature on “nominal” versus “real” bargaining power.

The next section provides examples from the political discourse in multiple contexts and its implications to broader phenomena. Section 3 provides the evidence from European national parliaments and 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.

---

8See [Warwick and Druckman (2001)] and [Frechette, Kagel, and Morelli (2005)] and the discussion on Section 4.4.
Motivating Examples and General Issues

This section discusses further examples of how the prescription that “the party with the most seats should form government” has been incorporated into the discourse of voters and politicians and its relation to a broader pattern where pluralities are reinterpreted as majorities (Maskin and Sen 2016).

Examples from New Zealand, the United Kingdom, Italy, and Spain. Since a 1994 electoral reform, no party in New Zealand obtained the majority needed to form a government, but for seven consecutive elections, the most represented party formed the government. The first exception occurred in 2017, when the second-placed Labour party formed a government. This was perceived as due to a “maverick” third-placed party “kingmaker” and met with confusion by voters. As put by the governor-general, “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.”

Similar issues appear in British politics. During campaigning for the 2015 general election, the leader of the Liberal-Democrat Party (Nicholas Clegg) stated that “the party that gets the most votes and most seats, in other words the party that gets the biggest mandate from the British people, even if it does not get a slam-dunk majority, it seems to me right to give that party the space and the time to try and settle a government” (Perraudin 2015) and Scottish Labour Party leader Jim Murphy 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, but fell short of a majority. The party argued that the most voted party appointing the prime minister would be “more democratic.” As one of its 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 agree with these statements. A nationally representative poll found that 55% of Spaniards agree that “it is more democratic that the most voted party forms the government, even if that party does not have an absolute majority of the votes” (El Pais 2015).10 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 Pais 2018). Subsection 4.2 discusses a related case study from Olivenza’s 2011 election.

Majorities versus pluralities. An interesting common aspect in the four contexts discussed above (and the Canadian example in the introduction) is that they appeal to notions of what is

---

9The New Zealand example is also illustrative since, in the seven elections between 1994 and 2017, two parties (Labour and National) alternated in power. Government formation in New Zealand has an “absence of formal procedures.” The description above and quotes are from Daly 2017. The comment section in Daly 2017 provides many examples of voters’ opinions (e.g., “clearly the highest polling party has the largest mandate to negotiate with a potential coalition partner”).

10One caveat is that the question does not clarify what disagreeing implies. Given the “more democratic” wording and “even without an absolute majority” qualifier, a 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.”

11Similar pledges have been made at regional governments, for example in Andalusia (Abr 2004).
“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 it could also be argued that a majority of voters did not vote for the party being labeled “the people’s choice.” Moreover, the Canadian case raises another interesting point: the sitting prime minister agrees that his incumbency status is less important than “winning the most seats” when it comes to appointing a prime minister.

These examples (as well as our empirical results) illustrate Maskin and Sen’s (2016) argument on the confusion (or reinterpretation) of pluralities as majorities discussed in the introduction. A 2004 campaign speech by PSOE leader José Luis R. Zapatero illustrates this connection. Zapatero promises to try to form a government only if his party obtains a “sufficient majority.” He clarifies 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.”

**Example of a broader pattern: Donald Trump’s nomination.** An example of how this issue applies to contexts beyond government formation is Trump’s nomination as a presidential candidate. The Republican Party’s rules state that a majority of delegates is needed to nominate a candidate. However, early in the 2016 primaries, many expected Trump to achieve a plurality, but not a majority, of delegate votes. Commentators (and Trump himself) declared that the candidate with the most delegate votes should be the nominee. 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.” 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

This section provides two results regarding government formation in European parliaments. First, most represented parties that have “just one more seat” than the second most represented are more likely to appoint prime ministers. Second, cases where the most represented party appoints the prime minister are associated with shorter delays in government formation and longer government durations.

**Data and context.** Our data covers 308 episodes of government formation following an election in 28 European countries in the 1944-2010 period. It contains the party affiliation of the appointed executive (prime minister) and the number of seats of each party in the lower house. In most cases,
there is no formal institutional advantage for the party with the most seats in forming government.\footnote{The only exception is Bulgaria, which in 1993 stipulated that the most voted party should be the first formateur. Repeating our analysis excluding Bulgaria leads to similar results (Appendix Figure \ref{fig:appx} and Table \ref{tab:appx}).} 
It is possible that some countries have informal rules, such as the head of government asking the most represented party to be the first formateur (the party which makes a proposal put to a vote). To the extent these are not coded in laws, they can be understood as norms. Moreover, \cite{Diermeier_2004} show that formateurs are not chosen by seat rank in a sample of European parliaments.\footnote{\cite{Diermeier_2004} show that formateur selection is better explained by selection probabilities being proportional to seat shares than following seat ranks. Their maximum likelihood estimation aims at fitting all the data, while we focus on behavior near a RDD cutoff. While Diermeier and Merlo’s (2004) sample includes only 11 countries, restricting our analysis to only the countries in their sample does not affect our results (Appendix Figure \ref{fig:appx} and Table \ref{tab:appx}).}

3.1 Effect of Having the Most Seats on Appointing Prime Ministers

\textbf{Empirical strategy.} We document this result by leveraging a RDD in which 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.\footnote{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 since a party must be labeled having the “most” and “second most” seats to enter equation (1).}

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 for party $p$ as:

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

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 party $p$ appointed the relevant executive member for the entire term following the election at $t$. We refer to this outcome as “appointing the prime minister (PM)” throughout. The effect of having the most seats is given by $\lim_{x_{pit} \downarrow 0} \mathbb{E}[y_{pit} | x_{pit}] - \lim_{x_{pit} \uparrow 0} \mathbb{E}[y_{pit} | x_{pit}]$, which can be estimated by a local polynomial regression:

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

using observations within a given bandwidth around the threshold. $g_0$ and $g_1$ are polynomials estimated separately on each side of the cutoff. Thus, $\theta_1$ captures the effect of having the most seats (instead of having the second most seats) in a parliament where the two parties with the most seats almost tied in seats. Baseline estimates use a linear specification ($g_0 = g_1 = x_{pit}$), as suggested by \cite{Lee_2010}, and the \cite{imbens} optimal bandwidth. We discuss the robustness to different bandwidths 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 be distributed symmetrically around 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
Second, conditioning which observations enter the sample by a variable that varies only at the election or country level will not affect the RDD’s “internal validity” (balance in covariates around the cutoff). For example, restricting the sample to only the cases where a party does not have a majority of seats cannot create imbalance 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 standard errors clustered at the country level.

Main result. Figure 1a plots the probability that a party appoints the PM against the seat share difference between the two parties with the most seats (parties placed third and lower are excluded from the sample). The 504 observations are aggregated into bins of 1-p.p. width of the running variable ($x_{pit}$) and 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 $\theta_0$ 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 PM (the estimated $\theta_1$). 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 linear specifications are robust to the choice of bandwidth (Figure A4). Using the full sample and a quadratic or cubic polynomial yields similar results, shown in columns (3) and (4).

Interestingly, this effect is not driven by an increased probability of being in the ruling coalition (having cabinet positions). Figure A5 replicates the exercise of Figure 1a 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.

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 have appointed the incumbent PM. Panel B of Table 1 present the analogous regression results.

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

---

18Figure A4 presents the distribution of observations, demonstrating this symmetry.

19Unfortunately, our data does not contain information on the share of cabinet positions each party obtained.

20Party 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 A6). If a particular “family” was more likely to be on a particular side of the cutoff, we would observe a non-zero effect.
third-placed party, while the second placed cannot. To investigate if such differences drive our results, Figure A7 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. The effects are strikingly similar to those from the main sample. This is consistent with the effect of “having the most seats” in itself, as opposed to higher bargaining power associated with the ability to form different coalitions.

To further explore this issue, we calculated the Shapley-Shubik power index for each of the 504 parties in the sample. Figure A8 repeats the exercise of Figure 1a using the index as the outcome. Parties in the 1-p.p. bin just left of the cutoff have an average index of 0.284, while those just right have essentially the same average (0.293). This suggests that the previous results are not driven by the differences in bargaining power captured by the index. Moreover, Figure A9 repeats the exercise of Figure 1a using the Shapley-Shubik index as the running variable, and the size of the “jump at the cutoff” is similar to the one in Figure 1a. Panel C of Table A1 presents the corresponding estimates.

Ideally, one could restrict the sample to cases where the top two parties have the same Shapley-Shubik index. However, this subsample would include only 46 elections, of which only six would involve the two most represented parties being one percent of seat share apart. This highlights the value of studying Spanish municipalities, where there are 2,898 cases with the two most voted parties tying in number of seats (implying they have the exact same Shapley-Shubik index).

### 3.2 Bargaining Delay and Government Duration

**Delays in government formation.** Delays in government formation can be costly, with previous literature associating delays in government formation with financial market volatility.\(^{22}\) They are also frequent. In our sample, it takes on average 38 days for a government to be formed when no party has a majority (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 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%.

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

\(^{21}\)The index is described in Shapley and Shubik (1954). Intuitively, the index measures the fraction of sequences in which the party is pivotal in forming a coalition if parties entered in coalitions sequentially. It varies from zero (the party is not pivotal in any sequence) to one (the party has a majority). We obtain similar results replicating the analysis using the Banzhaf III (1964) index (which measures the fraction of possible majority coalitions in which the party would be pivotal) and the Coleman-Shapley index (which combines features of the two previous indexes and is described in Casajus and Huettner (2018)). The correlation between these indexes in our data is always above 0.99.

\(^{22}\)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.” The quotes are from Golder (2010), which also surveys previous literature on the causes and consequences of bargaining delays in government formation. Figure A10 plots the cumulative distribution function of delays in government formation in our data for each of the subcases discussed.
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. 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 coefficient remains negative and
sizable.

The results suggest that the phenomenon we study may affect welfare. Another (more speculative)
interpretation is that it arises exactly to shorten such delays. It is important to caveat, however, that
results in Table 2 are based on correlations and, although robust to controls, it is difficult to 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 (number of days in power) as the outcome.\(^23\) If PMs from the party with the most seats
are perceived as a “more democratic” choice, they may 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 the previously discussed 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, the results from European national parliaments have the drawbacks
of a small sample size and the inability to fully control for potential differences in bargaining power
associated with additional seats. Given Spain’s numerous municipalities (and small council sizes), we
can observe many elections where the two most voted parties tie in seats.\(^24\)

4.1 Context, Data, and Empirical Strategy

**Electoral Rule and Government Formation Procedure.** Spanish national law regulates how
the formation of municipal governments. 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. Council size is a function of municipal population
(Table A2). Each party presents a ranked list of candidates. In the election, each voter picks one of

\(^{23}\) We approximate term length by the days between the election that led to that government and the next election.

\(^{24}\) To the best of our knowledge, there is no other context with a comparably large number of ties in seats. Contexts
where the executive is directly elected 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 covers more than 5,900 municipalities).
the party-lists. The conversion from party votes to seats follows D’Hondt rule.\footnote{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 $i$th row and $j$th column is the number of votes won by the $i$th 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). Political parties must also obtain at least 5\% of the votes to receive seats. This system refers to municipalities with more than 250 inhabitants. Other municipalities use an open-list system and are excluded from our sample. \cite{Sanz2017} and \cite{Gonzalez-EirasSanz2018} estimate the effects of the change in electoral rules at the 250-person cutoff on turnout and women’s representation, respectively.}

In their first meeting after the election, councilors choose a mayor (alcalde). Leaders of each party (the candidates ranked first in the lists) are eligible to be mayor. Each councilor votes for one of the leaders or abstain. If a leader has more than 50\% of votes, she is appointed mayor. If no leader 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.\footnote{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 \textit{general election}, then a lottery is run among the tied parties (in 1979 and 1983, such ties in \textit{general election} cases were decided by appointing the oldest party leader). Exact ties in votes in the \textit{general election} are very uncommon, and its few occurrences are deleted from our sample.}

The mayor can be replaced at any moment in the term by 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. Only one censure motion can occur per term. Another possibility 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., budgets). The number of votes required for the mayor to lose the motion depends on the context (e.g., in a budget vote, the mayor loses if nays outnumber yeas). If a mayor loses the motion, the council elects a new mayor following rules used for electing the mayor for the first time. There are no “off-schedule” general elections: citizens only vote every four years.\footnote{If a mayor resigns, is convicted of a crime, or dies, a new mayor is chosen by the same procedure. When appointing a new mayor after a removal, the candidate from the party of the removed mayor is the next person in the party-list.}

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.

**Importance of Municipal Governments.** Municipal governments manage approximately 15\% of the Spanish public expenditure (6\% of GDP). Spanish law dictates which services must be provided by municipal governments. They collect taxes on residential properties, businesses, vehicles, as well as fees and user charges. Total tax and fee revenue by municipalities 4\% of Spanish GDP. Hence, municipal-level policymaking is consequential to voters.\footnote{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).}

The mayor can be replaced at any moment in the term by 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. Only one censure motion can occur per term. Another possibility 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., budgets). The number of votes required for the mayor to lose the motion depends on the context (e.g., in a budget vote, the mayor loses if nays outnumber yeas). If a mayor loses the motion, the council elects a new mayor following rules used for electing the mayor for the first time. There are no “off-schedule” general elections: citizens only vote every four years.\footnote{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 \textit{general election}, then a lottery is run among the tied parties (in 1979 and 1983, such ties in \textit{general election} cases were decided by appointing the oldest party leader). Exact ties in votes in the \textit{general election} are very uncommon, and its few occurrences are deleted from our sample.}

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.

**Importance of Municipal Governments.** Municipal governments manage approximately 15\% of the Spanish public expenditure (6\% of GDP). Spanish law dictates which services must be provided by municipal governments. They collect taxes on residential properties, businesses, vehicles, as well as fees and user charges. Total tax and fee revenue by municipalities 4\% of Spanish GDP. Hence, municipal-level policymaking is consequential to voters.\footnote{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).}
have substantial control over the level allocation of expenditures (by preparing budgets). Spanish municipal governments exemplify 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 restoration of democracy. Elections have occurred in four-year intervals since 1979. The source is the Ministerio del Interior (Ministry of Internal Affairs). 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.

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 Ministry. Information about the allocation of cabinet positions within municipalities is also unavailable. Given that mayors may not necessarily serve the entire four-year term, we define our main outcome as a dummy equal to one if the mayor that spent three quarters of the term in power during the term belongs to that party.

Characteristics of Elections Close to the Cutoff. There are 438 elections where the two most voted parties tie in seats and their vote share difference is below 1 p.p.. In these cases, 90% have councils such that a majority requires support from any two of the three most voted parties. This includes both cases where only three parties received representation and cases where the fourth placed party cannot be pivotal for 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.”

In these 438 cases, councils are 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, with the third most voted obtaining vote (seat) share of 17.7% (16.7%).

Empirical Strategy. We implement a RDD in a 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). The implementation 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:

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

where \(v_{it}\) is the vote share difference between the first and second most voted parties. Thus, if \(x_{pit} > 0\),

---

29 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. Including the 1979 and 2015 data does not affect our results (Table A4).

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

31 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 similar 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).
party \( p \) has the most votes. The same three points discussed in Subsection \[\text{[65]}\] on the symmetry of variables that are constant across parties within an municipality-year, conditioning the sample on such variables, and clustering standard errors at the municipal level apply here.\[\text{[65]}\]

### 4.2 Effects of Being Most Voted

Figure 2a 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. The jump at the cutoff is the graphical counterpart to \( \theta_1 \). Figure 2a 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.\[\text{[65]}\]

A perhaps surprising pattern that an upward slope is not observed in Figure 2a. Conditional on the vote ranks, higher vote margins for the most voted parties are not associated with increased odds of appointing the mayor. Such slopes must be interpreted with caution, as there are both compositional effects and omitted variables that can affect them.\[\text{[65]}\]

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 \( \theta_0 \) from equation \[\text{[65]}\]). Column (1) indicates that the most voted that almost tied in seats is 19 p.p. more likely to appoint the mayor (\( \theta_1 \)). Both figures are based on using only the 2028 observations from elections where the top-two parties are only 2.32% of the total votes apart (the optimal bandwidth). Column (2) compares the average outcome for the first and second most voted parties that are only 1% of total votes apart and finds a similar effect. Even focusing on the 46 observations from close cases (bandwidth below 0.1%), the estimated effect is 0.522 (s.e. = 0.167). Results are robust to the choice of bandwidth (Figure A14). Using the full sample and a quadratic or cubic polynomial yields similar results, shown in columns (3) and (4).\[\text{[65]}\]

**Interpretation.** As discussed in the introduction, comparing the two most voted parties that obtained almost the same number of seats identifies the effect of being most voted. When comparing two groups that should have the same bargaining power, differences in outcomes isolate the effect of the most voted label. Note that an effect size of 20 p.p. is both consistent with 20% of the municipalities

\[\text{[65]}\] Figure A11 presents the distribution of observations, demonstrating the symmetric distribution.\[\text{[65]}\] 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.\[\text{[65]}\] Such compositional effects and “omitted variables” are constant when comparing cases around the cutoff. A possible compositional effect occurs since smaller councils are less likely to be further away from the cutoff (e.g., no two parties with a 8 p.p. vote share difference can have the same number of seats in a 21-member council, but that is possible in a 7-member council). However, Figure A12 replicates Figure 2a for specific council sizes and indicates the lack of slope is not driven by such compositional effect. Figure A13 replicates Figure 1a for specific council seat compositions.\[\text{[65]}\] Table A3 replicates Panel A of Table 4 when defining “appointing the mayor” in different manners. Panel E of Table A4 restricts the sample to the cases where two of the top three most voted parties are needed to form a majority. Table A5 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 4.
having a effect equal one and with all municipalities in the sample having an effect that only binds
with 20% probability (or a combination of these two extreme cases).

**Covariate Balance.** Figure 2b repeats the exercise from Figure 2a, but with the lagged outcome
on the y-axis. It 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 [A15] 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 [B] present the analogous
regression results. The point estimates are close to zero and statistically insignificant. These results
suggest that incumbent mayors or the major parties are unable to manipulate election results. As
previously discussed, for any variable that does not vary across parties within an election, there is
perfect balance by construction and there cannot be “bunching” of municipalities around the cutoff.

**Is the effect present when the second and third most voted are aligned?** We now focus
on cases where the right-wing PP and the left-wing PSOE are the most voted parties, and the left-wing
IU is the third most voted. The left-wing parties (PSOE and IU) thus have a combined majority are
able to appoint one of their leaders as mayor, regardless of whether the PP is the most voted.

The red triangles on Figure 3 repeat the exercise of Figure 1a, but restricting 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 [A5]
provides the corresponding table.

The surprising aspect is 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). 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 “most voted
party appoints the mayor” prescription “bites” even in the cases when aligned second and third most
voted parties can form a coalition to appoint a mayor of their own.

**An example.** The results of Olivenza’s 2011 election can illustrate our results. The PP obtained
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.

---

36 See Figure [A11] which shows how many observations are in each bin of Figure 2.
37 The estimates on Table [A2] are obtained by estimating equation (11) 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. The subsamples are not defined by which party is most voted and thus produce a balanced RDD.
Moreover, in a proportional representation system that does not formally reward the most voted, a majority of voters preferred the left-wing parties, but being most voted seemed to matter beyond that.

Appendix A discusses three additional sets of results. First, it discusses why the possible interpretation that mayoral appoints are symbolic is unlikely and provides results on deputy mayor appointments. Second, it shows that the results are likely driven both by coalitions where the most voted party forms a coalition with the second- and third-most voted. Third, it shows that the effect is not stronger for “powerful” parties that are in power at the national or regional level.

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.

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 parties to the right are the second most voted. A discontinuity at the cutoff is visible, although it is smaller and noisier than in Figure 2a. No discontinuity is visible in the placebo graph (Figure 3b) which plots lagged outcomes, increasing confidence that the jump observed in Figure 3a is not driven by noise.

Table 4 provides the regression results. Column (1) indicates that a third most voted party that almost ties with the second most voted 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.” Columns (2)-(4) indicate that effects of similar magnitude are estimated using different specifications. This result suggests that the effect generalizes to lower ranks. Since there is no status quo rule (or any other formal differential treatment) benefiting second versus third placed parties and thus it cannot play a role in the effects on Figure 3a and Table 4. Moreover, it suggests the similar effects of being 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.

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

38Our sample contains 1,565 elections (from 1,204 unique municipalities) satisfying these two conditions.
39Panels B and C and Figure A10 show the covariate balance and Figure A13 shows the number of observations in each bin of Figure 3a. Figure A12 presents the robustness of the estimates to bandwidth choice.
40Figure A17 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 these.
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. Whether an additional vote leads to one additional seat is defined by the rounding in D'Hondt rule. 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. It is useful to separate the cases where the most voted party obtains one more seat 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 on party has more seats than others, its ability to form coalitions is not larger (any two parties can form a majority). The Shapley-Shubik or Banzhaf index is the same for both top two parties in this case.

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 only needs one other party to form majority, the second most voted may need two. The Shapley-Shubik or Banzhaf index in this case differs for the top two parties.

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). In this case, the Shapley-Shubik or Banzhaf index is one for the most voted (and zero for the second most voted).

These three cases are directly observable and we can thus define three separate subsamples accordingly. Figure 5 provides the regression discontinuity plot for these different subsamples. The blue circles correspond to the case where the two most voted parties tie in number of seats and are thus exactly the same as in Figure 2a. 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 two cases is perhaps surprising. It suggests that the effect of “being most voted” is of similar magnitude as “being most voted and obtaining an extra seat.” The corresponding estimates are provided on Panel B of Table A6, 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.

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, representing case 2 above. Panel C of Table A6 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. Lastly,

---

41 For example, a 11-seat council where party vote shares are 42.0%, 41.9%, and 16.1% will have a 5-5-1 seat allocation. If the vote shares are the slightly different 41.0%, 40.9%, and 18.1%, the seat allocation is 5-4-2.

42 Note that we condition the sample definition to a variable defined at the municipality-year level and thus maintains the RDD “internal validity” as discussed in Subsections 3.1 and 4.1.

43 Of the 37,122 elections in our sample, 7.8% have the two most voted tying in seats (the focus of Subsection 4.2), and 7.9%, 8.6%, and 75.7% in cases (1), (2) and (3), respectively.
receiving a majority of seats (brown diamonds) makes a party almost surely appoint the mayor. Panel D of Table A10 provides the corresponding estimates.

These sets of results suggest that the effect of rank labels is comparable to the effects of additional seats, which have been the previous focus of 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. Our results suggest an additional mechanism that may help explain the puzzle.

5 Theoretical Framework

This section presents a model to illustrate a specific mechanism that can drive our results and yields predictions that guide the subsequent empirical analyses. Our starting point is a framework of political accountability (Barro 1973, Ferejohn 1986, Persson, Roland, and Tabellini 1997). We add to it legislative bargaining and 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 to appoint a mayor. However, parties’ representation is set at this point and they may bargain and form coalitions based on factors that ignore voters’ interests (e.g., rent allocations). This creates a conflict of interest between voters and parties and a reason for disciplining. The model has multiple equilibria, which can be interpreted as different 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. An alternative modeling strategy would be to assume voters inherently prefer a mayor from the most-voted party and will punish lower ranked parties that appoint mayors (or reward most-voted mayors). Doing so will naturally lead to outcomes similar to the equilibrium described in Proposition 2. Such a model could be interpreted as formalizing a norm as a principle of correct action that is internalized in preferences. In our model, norms are interpreted as equilibrium outcomes (self-enforcing mutual beliefs).

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 events occurring, each of them being better dealt with by a given mayor, or only one party in each period having an honest or competent leader.

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 (rents are indivisible).
The bargaining procedure is as follows. If one party received a majority of votes in the previous election, it can single-handedly choose the mayor. If no party had a majority of votes, then one party is randomly “recognized” (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 matches the one round of voting by majority rule feature of Spanish municipalities. 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. Parties’ recognition probabilities are a continuous function of vote shares.

Parties’ preferences and the bargaining procedure are thus independent of voters’ welfare and the states of the world. This creates a dissonance between voters’ and parties’ interests. Voters prefer the mayor that matches the state, but the choice of mayor may be determined by unrelated factors. If states were observable, voters would award a majority to the party matching the state.

**Uncertainty and information structure.** However, states of the world are not directly observed by voters or parties. The probability that state $s$ occurs in period $t$ is $p^s_t$, with $p_t$ denoting the vector $[p^A_t, p^B_t, p^C_t]$. Moreover, voters and parties face uncertainty about the vector $p_t$, which is drawn every period from a (common knowledge) distribution $G(p_t)$ that is independent and identically distributed over time. Each voter individually observes a signal $\sigma_t$ every period. The three possible signals are also $\{A, B, C\}$, with the probability the signal is $s_t$ given by $p^s_t$, drawn independently for each voter.

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 $G(p_t)$. After she observes her private signal of value $i$, she forms new beliefs $\Pr(s_t = k|\sigma_t = i)$ for all $k \in \{A, B, C\}$, which informs her vote. After observing the election results, she updates again, based on other voters’ strategies. If all citizens vote according to their signals (e.g., vote for $A$ if signal is $A$), she will expect the probability that the state is $s$ to be the vote share of party of type $s$. However, by this time parties’ representation in the legislature is already defined and, when another election occurs, a new vector $p_{t+1}$ and state $s_{t+1}$ will be drawn, making the previous information irrelevant.

We assume that $G(p_t)$ 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 $G(p_t)$ 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 $p_t$. Based on this vector it draws the state of the world and individual signals. Citizens then cast their votes. There are six possible votes: voting for one of the incumbent parties $A^I$, $B^I$, or $C^I$ that were in office in the preceding period or voting for one of challenger 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

---

40 It is analogous to parties having votes weighted by their previous election’s votes and deciding by majority rule.

41 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.

48 All realizations of $G(p)$ satisfy $p^A + p^B + p^C = 1$. 

18
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 as a way to discipline incumbents to behave in consonance with voters’ interests. The assumption of 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 could be more realistic, they would complicate the model and create a conflict between the information aggregation and disciplining roles of elections. Note, however, that many parties, with presence at the national, regional, and municipal level, operate in Spain. If the types of parties are interpreted as ideologies (e.g., left, center, right), this can be interpreted as multiple leftist parties (some regional or municipal) that can replace one another.\(^4\) While the model assumes a specific bargaining procedure, Appendix B discusses how the results are robust to allowing multiple rounds of bargaining and rents to be divisible across parties.

**Equilibria.** We restrict our attention to sequentially rational equilibria in which every voter chooses a pure strategy that conditions her decision on her last observed signal and the incumbent party’s behavior in the preceding period. All parties choose pure strategies that condition only on the result of that period’s election.

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. 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 (when others follow 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 Section 4. Those results, however, can be captured by the following equilibrium.

**Proposition 2.** If \(G(p_t)\) 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

\(^4\)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.
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, 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 described in Section 4. Even when the two most voted parties are 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 will appear to be based on ranks and bind behavior even in cases where the consequences are minimal (parties almost tying).

In the equilibrium described in Proposition 2, the most voted party always appoints the mayor. In practice, second placed parties appoint the mayor with non-trivial frequency. There are two ways to reconcile this with the model. The first is that not all municipalities are in the equilibrium with the norm. Under this interpretation, our effects pin down the share of municipalities following the norm.\(^5\) Another possibility is to incorporate deviations from the norm. These can occur due to trembling hand shocks in proposal strategies that make second placed parties appoint mayors on the equilibrium path and predict that voter punishment for such second placed parties can be observed.

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 continuing to follow it and perhaps obtain rents in a future period when it is the most voted party. Whether this condition is satisfied can be approximated empirically in two manners. First, we can observe if the third-placed party 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

---

\(^5\)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.
third most voted party has a larger vote share. The model predicts that the norm applies uniformly to all parties. This is consistent with the evidence we discuss in Section 4 and Appendix A: effects are similar regardless of whether the most voted party is aligned with higher levels of government or the third-placed party.

**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 does not directly address equilibrium selection (why some places adopt the norm), its optimality can provide an explanation for its prevalence.

6 Empirical Implications of the Theoretical Framework

The section above described three additional testable predictions from the model. First, voters punish second most voted parties that appoint mayors. Second, the effect of being most voted is larger when the third-placed party is stronger. Third, following the norm leads to politicians of better quality and higher welfare. This section provides evidence on the first two, while Appendix D addresses the third by studying the association between appointing most voted mayors and lower corruption.

6.1 Do Voters Punish Second-Placed Parties That Appoint Mayors?

We test whether second most voted parties that appoint the mayor go on to lose votes compared to first most voted parties that also do so. For identification purposes, we focus on 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.

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

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

where \(m = 1\) if party \(p\) appointed mayor, \(f = 1\) if party is first-place. The sample includes only first or second most voted parties and \(\gamma\) is the effect of appointing mayor for the second most voted and \(\gamma + \beta\) is the effect for the first most voted. 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 \(\alpha\). 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

51 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.

52 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.

53 This is the same sample that appears in column (2) of Table 3. The sample size differs since not all parties run in two consecutive elections, or local parties change names making it impossible to identify them over time.
parties appoint mayors. Columns (2) and (3) show similar results when a full set of region and party indicators is added (which control for region-time and party-time variation, given the first-difference specification).

**Dynamics and Parallel Trends.** To test 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 6a provides the event-study counterpart for equation (2) from estimating

\[
(v_{pi,t+k} - v_{pi,t}) = \alpha_k + \beta_k m_{pit} + \gamma_k f_{pit} + \delta_k f_{pit} + \epsilon_{pit}
\]

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

**Interpretation.** Given the triple-difference nature of the estimation, it is not clear whether the effects of Figure 6a 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). Figure 6b illustrates this by providing the double-difference event study graph for both second placed parties and first placed parties separately. In particular we estimate equation 4 separately for only second most voted parties (red squares) and first most voted parties (blue circles). 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). First most voted parties are on a similar trajectory before appointing a mayor, but go on to gain even more after.

While Figure 6a indicates a non-zero effect four years after the mayoral appointment, it is not 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

Table 6 provides evidence supporting the model’s prediction that, in elections where the third most voted party has a higher vote share, the effect of being most voted should be stronger. It repeats our estimation of the main results (Table 3), but separating the sample into cases where the third most voted party vote share is above (Panel A) and below (Panel B) the median.

---

54This 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.

55We 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%).
in Panel A, and it is possible to reject that the effects in both subsamples are the same in all specifications across columns (at the 5% level). Figure A18 provides the graphical counterpart.\footnote{Figure A18 is constructed similarly to Figure 1a, but for each subsample. Figure A18b 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 A18 shows that this heterogeneity holds within some council sizes.}

Table A7 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 A19 provides the graphical counterpart.\footnote{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.}

Lastly, we test whether voters’ punishment for second most voted parties that appoint mayors is more evident 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 7 estimates equation (2) for the subsample 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 differential effect for most voted parties ($\beta$) is larger in the former case.

7 Alternative Explanations

While the results we document are consistent with the existence of the norm modeled in Section 5, there are other possible explanations. This section discusses two alternatives (the status quo rule and voters preferring “winners”), while Appendix F discusses the role for politician’s outcome bias, parties making informal agreements, and the central government preferring most voted parties.

**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. 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 E discusses in further detail other reasons why we believe the status quo rule does not play a role in explaining our results.

**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 occur because “winning” attaches symbolic value to a party. This would in turn imply voters punish mayors from second most voted parties and tilt bargaining outcomes towards the most voted party.

First, note that this “alternative” explanation is, in some 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. This highlights the complementary character of the 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. 

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.” 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.

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 support this most voted advantage even when a natural alternative, such as programmatically aligned lower-ranked parties forming a winning coalition, is available.

We believe there are two fruitful avenues for further research. One is studying the welfare consequences of this effect. Another is investigating why this effect is present in some jurisdictions and not others. In light of our model, that would require understanding the equilibrium (norm) selection mechanisms.

References


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.


El Pa´ıs (2004). Zapatero anuncia que solo intentar´a formar gobierno si logra m´as votos que Rajoy. [https://elpais.com/diario/2004/01/19/espana/1074466803s50215.html](https://elpais.com/diario/2004/01/19/espana/1074466803s50215.html).

El Pa´ıs (2015). Clima social tras las elecciones del 24-M. [http://elpais.com/elpais/2015/06/19/media/14373713%2538.html](http://elpais.com/elpais/2015/06/19/media/14373713%2538.html).


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.
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 3a](image)

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 3b: Placebo Test: “Effect” of Second Most Voted on Lagged Outcome

![Figure 3b](image)
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.
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 B show how the share of votes for a second-placed (first-placed) party that appoints a mayor at \( t = 0 \) evolves relative to a second-placed (first-placed) party that does not, obtained by estimating \( \gamma \) and \( \gamma + \beta \) from equation (2) with different time horizons (see text for further details). Blue circles in Panel A represent the triple-difference event study: the difference between Panel B markers, obtained by estimating \( \beta \) from equation (4) with different time horizons (see text for further details).
Table 1: National Parliaments Data: Effect of Having Most Seats on Appointing Prime Minister

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>2nd-pl. Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Main Outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.202</td>
<td>0.303**</td>
<td>0.583**</td>
<td>0.387***</td>
<td>0.387**</td>
</tr>
<tr>
<td>Prime Minister</td>
<td>(0.130)</td>
<td>(0.210)</td>
<td>(0.122)</td>
<td>(0.145)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>224</td>
<td>24</td>
<td>504</td>
<td>504</td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: Lagged Outcome (Placebo Test)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.414</td>
<td>-0.0316</td>
<td>0.167</td>
<td>0.0852</td>
<td>0.136</td>
</tr>
<tr>
<td>Prime Minister, t – 1</td>
<td>(0.163)</td>
<td>(0.332)</td>
<td>(0.122)</td>
<td>(0.129)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>152</td>
<td>24</td>
<td>504</td>
<td>504</td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: Covariate Balance (Outcome Predicted from Party Ideology)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.417</td>
<td>-0.021</td>
<td>-0.035</td>
<td>-0.015</td>
<td>-0.009</td>
</tr>
<tr>
<td>PM (Predicted)</td>
<td>(0.028)</td>
<td>(0.078)</td>
<td>(0.026)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>222</td>
<td>24</td>
<td>504</td>
<td>504</td>
<td></td>
</tr>
</tbody>
</table>

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

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

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 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 of the classes of seat distributions in a legislature described by [Layer 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

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>2nd-pl. Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Main Outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.353</td>
<td>0.185***</td>
<td>0.203***</td>
<td>0.295***</td>
<td>0.241***</td>
</tr>
<tr>
<td>Mayor</td>
<td></td>
<td>(0.059)</td>
<td>(0.044)</td>
<td>(0.037)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>N</td>
<td>2028</td>
<td>876</td>
<td>5796</td>
<td>5796</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel B: Lagged Outcome (Placebo Test)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Party Appointed</td>
</tr>
<tr>
<td>Mayor, $t - 1$</td>
</tr>
<tr>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel C: Covariate Balance</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Party is PP</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Party is PSOE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>N</td>
</tr>
</tbody>
</table>

Specification: Linear Means Quad. Cubic
Bandwidth: Optimal <1% 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

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

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

<table>
<thead>
<tr>
<th></th>
<th>Outcome $(v_{pm,t+1} - v_{pm,t})$</th>
<th>Lagged Outcome $(v_{pm,t} - v_{pm,t-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Mayor$_t$*Most Voted$_t$ ($\beta$)</td>
<td>4.834**</td>
<td>4.232**</td>
</tr>
<tr>
<td></td>
<td>(1.931)</td>
<td>(1.881)</td>
</tr>
<tr>
<td>Mayor$_t$ ($\gamma$)</td>
<td>2.868**</td>
<td>3.091**</td>
</tr>
<tr>
<td></td>
<td>(1.449)</td>
<td>(1.469)</td>
</tr>
<tr>
<td>Most Voted$_t$ ($\delta$)</td>
<td>-1.693</td>
<td>-1.532</td>
</tr>
<tr>
<td></td>
<td>(1.393)</td>
<td>(1.355)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.160</td>
<td>-0.681</td>
</tr>
<tr>
<td></td>
<td>(0.785)</td>
<td>(3.025)</td>
</tr>
<tr>
<td>Region FE</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Party FE</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Only elections w. vote</td>
<td></td>
<td></td>
</tr>
<tr>
<td>share of 3rd &gt; median</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Only elections w. vote</td>
<td></td>
<td></td>
</tr>
<tr>
<td>share of 3rd &lt; median</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>664</td>
<td>664</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>2nd-pl. Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Third most voted party vote share above median</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.290</td>
<td>0.306***</td>
<td>0.290***</td>
<td>0.364***</td>
<td>0.334***</td>
</tr>
<tr>
<td>Mayor</td>
<td></td>
<td>(0.078)</td>
<td>(0.056)</td>
<td>(0.048)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>N</td>
<td>1014</td>
<td>468</td>
<td>2756</td>
<td>2756</td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: Third most voted party vote share below median</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.430</td>
<td>0.0402</td>
<td>0.103</td>
<td>0.214***</td>
<td>0.118*</td>
</tr>
<tr>
<td>Mayor</td>
<td></td>
<td>(0.088)</td>
<td>(0.066)</td>
<td>(0.059)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>N</td>
<td>1014</td>
<td>408</td>
<td>3040</td>
<td>3040</td>
<td></td>
</tr>
</tbody>
</table>

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).
Appendix A: Additional Results on Spanish Mayoral Appointments

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 results would arise if mayoral appointments were 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 A12, however, indicates that RDD “jumps” are larger in municipalities with larger councils, which are also those with more inhabitants (Table A2). Fifth, it does not appear that mayors share power with other parties.

To further shed light on whether the mayoral appointments are just “symbolic,” we can study 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. 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).

---

1There 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.

2If 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).
Figure A21 repeats the exercise of Figure 2a, 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 A8 provides the corresponding estimates. The estimated effects are all positive and sizable. Hence, being the most voted party increases the chance of appointing the mayor and the share (or having all) of the deputy mayors. This is the opposite of what the hypothesis that the two most voted parties obtain equal amount of power would predict.

Which types of coalitions drive the effect? While we cannot directly observe which parties are supporting the mayor, we can indirectly study whether coalitions by the most voted and second most voted, or between the most voted and third-placed party, drives the effects. Although we cannot observe the supporting coalitions, we leverage previously discussed evidence that left-wing PSOE and the right-wing PP are unlikely to support each other’s governments. Figure A22 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 the effect is not driven by coalitions where the second most voted (or third placed) party supports the first most voted.

Is the most voted effect different for powerful parties? Figure A22 suggests similar effects on a party, regardless of whether it is one of the two dominant parties is Spanish politics or not. To further probe possible differential effects given how powerful a party is, Figure A23 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.

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

\[3\]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 A8 are smaller than in Table B.
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 $p_t = [p_t^A, p_t^B, p_t^C]$ denotes the probabilities of state $s_t \in \{A, B, C\}$ occurring. In particular, assume that $G(p_t)$ is such that its possible realizations are i) $p_t = [0.45, 0.35, 0.20]$, ii) $p_t = [0.35, 0.45, 0.20]$, and iii) $p_t = [0.20, 0.20, 0.60]$. Each realization can occur with probability equal to 1/3.

Hence, when a period starts, voters have priors that each of the states of the world are equally likely. After observing a signal $\sigma$ equal to $A$, a voter updates and then believes that the probability that realization (i) occurred is $\frac{0.45\cdot(1/3)\cdot(0.45+0.35+0.20)}{1/3} = 0.45$. 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 $p$ 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 $G(p_t)$ 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 $G(p_t)$ 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.\footnote{A proposer keeps $y$ and offers one randomly drawn other party $1 - y$. For the other party to accept, its payoff must be $1 - y > V$, where $V$ is the continuation value of this legislative bargaining game. The proposer thus optimally makes this inequality bind, so the proposal is accepted. Hence, the continuation value equals $V = \frac{1}{3}y + \frac{2}{3}(1 - y) = \frac{1}{3}$ and $y = \frac{2}{3}$.}

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

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

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

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: Do Municipalities that Appoint Most Voted Mayors Have Less Corruption?**

This appendix addresses the model’s prediction that municipalities following the norm have mayors of higher (state-specific) quality and thus higher voter welfare. In particular, we report an association between a municipality’s tendency to appoint most voted mayors and lower government corruption.

**Interpretation.** A formal interpretation of these results in light of the model can be obtained by assuming that different states of the world represent that one party \((A, B, \text{ or } C)\) is fielding the only honest candidate (or one party has no opportunity to engage in corruption) in that period. If voters prefer honesty (or less corruption) and have imperfect information about which party is more honest (observing only individual signals), the setup of the model extends to the most voted candidate being more honest and municipalities under the equilibrium from Proposition 1 (no norm) having more corruption than those under the equilibrium from Proposition 2 (norm).

The assumption that voters prefer less corruption seems plausible in the context of Spanish municipalities. Instances of municipality-level corruption lower trust in government (Solé-Ollé and Sorribas-Navarro, forthcoming) and incumbent’s votes (Costas-Pérez, Solé-Ollé, and Sorribas-Navarro [2012]).

An important caveat is that voters preferences over mayoral appointments (and having a most-voted mayor) may be driven mostly by considerations separate from corruption (e.g., how mayors allocate spending). However, it is not possible to observe voter welfare or to infer the quality of policymaking from observable policies (e.g., budgets or spending) or outcomes that governments have little control over (e.g., income or population growth). Thus, our focus on corruption is also partly driven by these data availability concerns.

Another caveat is that our model provides little guidance over which municipalities will adopt the equilibrium following the norm. Our model thus does not suggest an identification strategy for the effects of the norm on outcomes such as corruption. Hence, the results in this appendix are based on correlations and, although they appear robust to controlling for many relevant factors, it is difficult to rule out the possibility of other confounding factors or reverse causality.

**Context and data.** Government corruption in 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).

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 \((\text{corruption}_{it})\) 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.\footnotetext{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 information 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.}

Whether a municipality is acting consistently with the norm from Proposition 2 is not directly observable, but we can construct a proxy variable ($\text{norm}_i$) 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 $\text{norm}_i$ as missing, otherwise, we code $\text{norm}_i$ 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.

**Results.** To estimate the effect of the norm on corruption, we regress $\text{corruption}_{it}$ on $\text{norm}_i$, 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, $\text{norm}_i$ does not vary across time within a municipality.\footnotetext{This implies that we cannot include municipality fixed effects. The sample is an unbalanced panel of 2,390 observations from 450 unique municipalities. $\text{norm}_i$ is equal to zero 41% 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 $\text{norm}_i$ and thus can only generate attenuation bias.}

Column (1) in Table A9 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 $\text{norm}_i$ 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 $\text{norm}_i = 1$ 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 A9'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.

Lastly, we reiterate that the results reported in this appendix should be interpreted as a correlation between municipalities acting consistently with a “most voted party appoints the mayor” prescription and lower instances of corruption. Fully ruling out the possibility of omitted variables or reverse causality driving the result would require an identification strategy leveraging an understanding of why some municipalities act as if following the norm that arises in our model, which is beyond the scope of this study.
Appendix E: 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 A13).

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 \((A, B, \text{ and } C)\), indexed by \(i\). As in Spanish municipal councils, each party has only one candidate for mayor, which we also label \(A, B, \text{ and } C\). 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 \((B \text{ and } C)\) are ideologically aligned. Each party prefers to appoint the mayor itself. For parties \(B \text{ and } C\), 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 \(A \text{ and } B\) are tied in seats and \(C\) has the same or fewer seats than \(A \text{ and } B\), 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 \((a, a, a), (a, \phi, a), (a, a, \phi), (a, \phi, \phi), (a, b, b), (a, c, c), (b, b, b), (c, c, c), (\phi, a, a), (\phi, \phi, a), (\phi, a, \phi), (\phi, \phi, \phi), (\phi, b, b), \text{ and } (\phi, c, c)\). Out of these 14 Nash equilibria, \(A\) appoints the mayor in eight, \(B\) in three, and \(C\) in three.

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

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

Second, note that actions \(b\) and \(c\) are weakly dominated (by \(A\) and \(\phi\)) for party \(A\). Additionally, \(a\) and \(\phi\) are dominated by \(b\) (for party \(B\)) and \(c\) (for party \(C\)). Hence, there are only four Nash equilibria where a party is not playing a weakly dominated strategy: \((a, b, b), (a, c, c), (\phi, b, b), \text{ and } (\phi, c, c)\). In none of them, party \(A\) appoints the mayor, even though it is the status quo. Assuming parties choose a weakly dominated action in a small voting game is unattractive. For example, trembling hand deviation.

\(^{7}\)If the other two parties are not casting the same vote, \(A\) and \(\phi\) can guarantee the best outcome for \(A\). If only one other party is playing \(a\) or \(\phi\), those actions are strictly better than \(b\) and \(c\) for \(A\). If the other two parties are both voting \(b\) or \(c\), \(A\) is indifferent between all actions.

\(^{8}\)If both parties are not casting the same vote, any other action is strictly better. 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 \(b, a, \text{ and } \phi\). A similar argument applies to party \(C\).
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.

**Appendix F: Further Alternative Explanations**

**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.

**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. Panel A of Figure A20 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. 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.

---

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

10 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.
Lastly, note that discussion above regards parties making agreements at the municipal level. In principle, one could envision parties 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 A22 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.

**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 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.
Appendix Figures and Tables

Figure A1: Effect of Having the Most Seats, Excluding Bulgaria - National Parliaments Data

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. Bulgaria is excluded from the sample.

Figure A2: Effect of Having the Most Seats, Diermeier and Merlo (2004)’s Sample - National Parliaments Data

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. 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. The sample is restricted to the 11 countries studied by Diermeier and Merlo (2004): Belgium, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Netherlands, Norway, and Sweden.
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 A4: 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 A5: 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 A6: 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.
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.

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 the Shapley-Shubik index. 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 A9: Effect of Having the Most Seats, Shapley-Shubik Index as Running Variable - National Parliaments Data

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 the Shapley-Shubik index 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. Averages are calculated within 1 p.p.-wide bins (horizontal axis). Continuous lines are a quadratic fit over the original (unbinned) data.

Figure A10: 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 A11: 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.
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 A13: Effect of First-Place, by Council Type

Notes are same to those on Figure A12, 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 A14: 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.
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 A17: 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 A18: 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 A19: 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 the 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 A20: 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 A21: 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.
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.

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.
Table A1: Additional Results on Effect of Having Most Seats on Appointing Prime Minister

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>2nd-pl. Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
</table>

**Panel A: Excludes Bulgaria from Sample**

<table>
<thead>
<tr>
<th>Party Appointed</th>
<th>0.205</th>
<th>0.318**</th>
<th>0.583**</th>
<th>0.392***</th>
<th>0.397**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prime Minister</td>
<td>(0.131)</td>
<td>(0.210)</td>
<td>(0.124)</td>
<td>(0.147)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>220</td>
<td>24</td>
<td>496</td>
<td>496</td>
<td></td>
</tr>
</tbody>
</table>

**Panel B: Sample Restricted Countries in Diermeier and Merlo (2004)’s Sample**

<table>
<thead>
<tr>
<th>Party Appointed</th>
<th>0.214</th>
<th>0.300**</th>
<th>0.625**</th>
<th>0.250</th>
<th>0.414**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prime Minister</td>
<td>(0.127)</td>
<td>(0.205)</td>
<td>(0.138)</td>
<td>(0.155)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>132</td>
<td>16</td>
<td>340</td>
<td>340</td>
<td></td>
</tr>
</tbody>
</table>

**Panel C: Shapley-Shubik Index as Running Variable**

<table>
<thead>
<tr>
<th>Party Appointed</th>
<th>0.181</th>
<th>0.431***</th>
<th>0.560***</th>
<th>0.373***</th>
<th>0.411***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prime Minister</td>
<td>(0.138)</td>
<td>(0.0498)</td>
<td>(0.125)</td>
<td>(0.143)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>254</td>
<td>504</td>
<td>504</td>
<td>504</td>
<td></td>
</tr>
</tbody>
</table>

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 7.38%, 7.43%, and 0.183, for the three panels, respectively.
Table A2: Distribution of Council Sizes

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


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 A3: Effect of Being First (Instead of Second) Most Voted: Alternative Specifications

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

Specification: Linear Means Quad. Cubic
Bandwidth: Optimal <1% 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 First (Instead of Second) Most Voted: Including 1979 and 2015 Elections**

<table>
<thead>
<tr>
<th>Dependents Variable</th>
<th>2nd-pl. Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Main Outcome (mayor serves at least 3/4 of term), with 1979 data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.343</td>
<td>0.209***</td>
<td>0.204***</td>
<td>0.290***</td>
<td>0.242***</td>
</tr>
<tr>
<td></td>
<td>(0.0621)</td>
<td>(0.0417)</td>
<td>(0.0347)</td>
<td>(0.0437)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1882</td>
<td>980</td>
<td>6474</td>
<td>6474</td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: Outcome is appointing mayor for entire term, with 1979 data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.340</td>
<td>0.205***</td>
<td>0.218***</td>
<td>0.300***</td>
<td>0.254***</td>
</tr>
<tr>
<td></td>
<td>(0.0551)</td>
<td>(0.0405)</td>
<td>(0.0339)</td>
<td>(0.0427)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2206</td>
<td>980</td>
<td>6474</td>
<td>6474</td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: Outcome is appointing mayor for longer than other parties, with 1979 data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.366</td>
<td>0.223***</td>
<td>0.220***</td>
<td>0.304***</td>
<td>0.267***</td>
</tr>
<tr>
<td></td>
<td>(0.0610)</td>
<td>(0.0427)</td>
<td>(0.0351)</td>
<td>(0.0447)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2050</td>
<td>980</td>
<td>6474</td>
<td>6474</td>
<td></td>
</tr>
<tr>
<td><strong>Panel D: Outcome is appointing initial mayor, with 1979 data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.340</td>
<td>0.244***</td>
<td>0.245***</td>
<td>0.335***</td>
<td>0.288***</td>
</tr>
<tr>
<td></td>
<td>(0.0565)</td>
<td>(0.0424)</td>
<td>(0.0352)</td>
<td>(0.0445)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2282</td>
<td>980</td>
<td>6474</td>
<td>6474</td>
<td></td>
</tr>
<tr>
<td><strong>Panel E: Outcome is appointing initial mayor, with 1979 and 2015 data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.360</td>
<td>0.222***</td>
<td>0.233***</td>
<td>0.310***</td>
<td>0.267***</td>
</tr>
<tr>
<td></td>
<td>(0.0543)</td>
<td>(0.0409)</td>
<td>(0.0341)</td>
<td>(0.0431)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2464</td>
<td>1074</td>
<td>7004</td>
<td>7004</td>
<td></td>
</tr>
<tr>
<td><strong>Panel F: Main outcome, sample restricted to cases where “two parties out of top-3 needed for majority”, with 1979 data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.357</td>
<td>0.193***</td>
<td>0.205***</td>
<td>0.295***</td>
<td>0.244***</td>
</tr>
<tr>
<td></td>
<td>(0.0647)</td>
<td>(0.0443)</td>
<td>(0.0364)</td>
<td>(0.0459)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1790</td>
<td>888</td>
<td>6110</td>
<td>6110</td>
<td></td>
</tr>
</tbody>
</table>

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 A5: Effect of Being Most Voted: Cases with a Left-Wing Majority

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>2nd-pl. Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Effect for PSOE (conditional IU being third most voted)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSOE Appointed</td>
<td>0.543</td>
<td>0.267*</td>
<td>0.248**</td>
<td>0.417***</td>
<td>0.264*</td>
</tr>
<tr>
<td>Mayor</td>
<td>(0.153)</td>
<td>(0.118)</td>
<td>(0.109)</td>
<td>(0.145)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>155</td>
<td>64</td>
<td>423</td>
<td>423</td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: Effect for PP (conditional IU being third most voted)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP Appointed</td>
<td>0.102</td>
<td>0.242*</td>
<td>0.231**</td>
<td>0.313***</td>
<td>0.246*</td>
</tr>
<tr>
<td>Mayor</td>
<td>(0.146)</td>
<td>(0.110)</td>
<td>(0.101)</td>
<td>(0.139)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>155</td>
<td>64</td>
<td>423</td>
<td>423</td>
<td></td>
</tr>
</tbody>
</table>

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

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

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 focuses 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 A7: Effect of Being Most Voted on Appointing the Mayor, by Probability of Third-Placed Becoming the Most Voted

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>2nd-pl. Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel A: Third most voted party vote is the most voted at some election (t-3, t+3)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party Appointed</td>
<td>0.170 **</td>
<td>0.494***</td>
<td>0.396***</td>
<td>0.351***</td>
<td>0.453***</td>
</tr>
<tr>
<td></td>
<td>(0.152)</td>
<td>(0.111)</td>
<td>(0.117)</td>
<td>(0.153)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>218</td>
<td>96</td>
<td>550</td>
<td>550</td>
<td></td>
</tr>
</tbody>
</table>

| **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 A8: Effect of Being Most Voted on Deputy Mayors’ Allocation

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

Specification: Linear Means Quad. Cubic
Bandwidth: Optimal <1% Full Full

Standard errors clustered at the municipality level in parentheses. The unit of observation is a party-municipality-year. The sample is restricted to the two most voted parties in elections in which they tied in seats. Each figure in columns (1)-(4) reports a separate local polynomial regression estimate with the specified bandwidth and polynomial order. Separate polynomials are fitted on each side of the threshold. 2nd-Place Mean is the estimated value of the dependent variable for a 2nd most voted party that tied with the most voted party (using the specification in column 1). Optimal bandwidths are based on Imbens and Kalyanaraman (2012), being equal to 2.28%, 3.47%, 2.37%, and 3.30% for the four panels, respectively.
Table A9: Municipalities That Appoint Most-Voted Mayors Have Fewer Instances of Corruption

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm(_i)</td>
<td>-0.0253**</td>
<td>-0.0264**</td>
<td>-0.0259**</td>
<td>-0.0253**</td>
<td>-0.0209**</td>
</tr>
<tr>
<td></td>
<td>(0.0117)</td>
<td>(0.0109)</td>
<td>(0.0115)</td>
<td>(0.0118)</td>
<td>(0.0103)</td>
</tr>
<tr>
<td>N</td>
<td>2390</td>
<td>2390</td>
<td>2390</td>
<td>2390</td>
<td>2390</td>
</tr>
<tr>
<td>Dep. Variable Mean</td>
<td>0.0693</td>
<td>0.0693</td>
<td>0.0693</td>
<td>0.0693</td>
<td>0.0693</td>
</tr>
</tbody>
</table>

for \( norm_i = 0 \) cases

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).