# Electoral Institutions and Intraparty Cohesion*

Konstantinos Matakos† Riikka Savolainen‡ Orestis Troumpounis§
Janne Tukiainen¶ Dimitrios Xefteris‖

May 6, 2024

**Abstract**

By utilizing unique data capturing candidates' ideological positions in Finnish municipal elections and leveraging exogenous changes in council size at different population thresholds as a proxy for electoral rule disproportionality and the expected advantage to the election winner, we identify a positive effect of council size on party cohesion. We propose the following mechanism: if a more diverse set of candidates is electorally appealing but less efficient in serving policy-related goals, parties face weaker incentives to maintain cohesion in institutional settings, such as smaller councils, which reward higher vote shares more generously.

**Keywords:** Electoral systems; ideological heterogeneity; party cohesion; proportional representation; regression discontinuity design

**JEL codes:** C21; C72; D02; D72

*We are grateful to James Adams, Manuel Bagues, Peter Buisseret, Alessandra Casella, John Duggan, Olle Folke, Alexander Fouirnaies, Anthony Fowler, Bernard Grofman, Tasos Kalandrakis, David Kang, Eva Mörk, Tuomas Pekkarinen, Carlo Prato, Johanna Rickne, Jim Snyder, Stephane Wolton, the editor and anonymous reviewers for their constructive comments and suggestions. For valuable feedback we thank participants in the APSA, CRETE, EEA, EPSA, MPSA, PEPS, SIEP conferences, as well as seminar audiences in Bocconi, Chicago Harris, Columbia, Harvard, HECER, King's College London, Lancaster, Leicester, LSE, Padova, VATT and Zurich. This research is funded by the European Union (Tukiainen, ERC, INTRAPOL, 101045239). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. Edited by Anna Dreber Almenberg.

†King's College London, Strand campus, London, WC2R 2LS, UK

‡Swansea University, Bay Campus, Fabian Way, Swansea SA1 8EN, United Kingdom.

§University of Padova, via del Santo 33, 35123 Padova, Italy and Lancaster University, Bailrigg, Lancaster, LA1 4YX, United Kingdom.

¶*Corresponding author:* Turku School of Economics, Rehtorinpellonkatu 3, FI-20014 University of Turku; and VATT Institute for Economic Research, Arkadiankatu 7, Helsinki FI-00101. Email: janne.tukiainen@utu.fi

‖University of Cyprus, PO Box 20537, 1678 Nicosia, Cyprus.

# 1 Introduction

Parties and their elites control candidate selection and play an important role in shaping the ideological composition of legislatures and policy outcomes. In many cases, such as in Europe and most parliamentary systems, parties hold nearly complete authority to decide the composition of their candidate lists. However, our understanding of the determinants of candidate selection is still quite limited (see, for example, the conclusion in Dal Bò and Finan 2018 and the references therein). This paper focuses on electoral rule disproportionality as a determinant of candidate selection, empirically examining its relationship with parties' incentives to nominate a more or less cohesive list of candidates.

We refer to electoral rule disproportionality as the expected distortions generated by the electoral rule in favor of the election winner (Taagepera 1986). In proportional rules, a party's seat share in the elected body (e.g., council or parliament) is in expectation (almost) identical to its vote share. Conversely, disproportional electoral rules generate distortions in the mapping of vote shares to seat shares that in expectation favor the election winner. These distortions can be attributed to various specific institutional characteristics, among which the size of the elected body (Lijphart 1995; Herron et al. 2018). Indeed, in (very) large bodies, a party's seat share in the elected body is, in expectation, (almost) identical to its vote share. That is, the electoral rule is proportional. Conversely, in small bodies, distortions are expected to advantage the election winner, making the electoral rule more disproportional. How does the magnitude of such expected advantage for the election winner affect parties' choices when nominating their lists?

We inform our empirical approach by relying on a well-grounded conceptual framework where parties present a list of candidates *"tethered by a rubber band to the ideology espoused by the parties whose label they run on"* (Grofman 2008). On the one hand, parties have incentives to obtain high vote shares so that they can influence the implemented policy in their preferred direction. Parties can attain this by adopting a more pluralistic approach, akin to a catch-all party. This involves including a diverse set of candidates in their lists, thus attracting a broader spectrum of voters with diverse ideological leanings (Kirchheimer 1966). On the other hand, electing candidates that adhere to a variety of ideologies can be harmful for the party since they may promote policies far from the party's ideology. Under this tradeoff, what is the optimal degree of a party's ideological cohesion? Given that disproportional electoral rules are expected to favor the election winner, they strengthen parties' incentives to become electoral appealing and hence propose a rather heterogeneous list of candidates. In contrast, relatively proportional electoral rules offer weaker incentives for an increased vote share, leading to the formation of ideologically cohesive lists.

Our empirical analysis focuses on Finnish municipal elections. This setup offers the opportunity to establish a meaningful identification strategy and utilize a novel dataset that records individual candidates' policy positions. Causal evidence is obtained by focusing on quasi-experimental empirical evidence due to municipalities' council size being determined solely as a step-function of their population. That is, we use changes in the council size as a proxy for changes in the rule's disproportionality, allowing for a regression discontinuity design (RDD). Data on parties' ideological cohesion is obtained by leveraging on a unique dataset recording the individual candidates' policy positions in 2008 and 2012. These data come from the voting aid application of the Finnish public broadcasting company, YLE, and are further linked to electoral data and other candidate-level information. The questions in the voting aid applications allow us to construct two cohesion indices, one using all available questions, and one focusing on candidates' redistribution preferences.

We start our empirical analysis by showing that council size serves as a reliable proxy for changes in the rule's disproportionality (Section 3.4). That is, our first set of results validates our proxy by showing that, within our dataset, realized distortions in favor of parties securing high vote shares are more pronounced in small councils than in large councils. Consequently, this implies that the distortions imposed by the rule in our setting decrease with council size. The magnitude of the effect of crossing the threshold is a decrease of approximately 21% - 26% relative to the mean value of the outcome, or 40% - 67% of the outcome standard deviation, depending on the measure of distortions. While this result aligns with theoretical predictions and is not the primary focus of our study, this paper, to the best of our knowledge, is the first to offer causal evidence regarding the impact of council size on realized distortions. Indeed, within the existing literature, this relationship has been regarded as a folk theorem, with supporting arguments primarily being theoretical or based solely on correlations (Benoit 2000; Herron et al. 2018).

Our main results in Section 3.5 show that, in the elections for smaller councils, competing lists tend to be less ideologically cohesive than in the elections for larger councils. The magnitude of the effect of crossing the threshold translates into a decrease in our two cohesion indices by 7% relative to the mean value of these outcomes, or 20% - 30% of the outcome standard deviation.

As our RDD identifies only the total effect on cohesion, we use our rich data to perform covariate balance tests and do not observe evidence of other jumps. For instance, our empirical setting permits us to focus on the effect of council size on intraparty cohesion without worrying for potential confounding effects through a change in the number of parties or candidates (Rae 1967; Taagepera and Shugart 1989; Cox 1990). Furthermore, no other policy changes take place at the thresholds that determine the council size. We also show that our results are not consistent with an alternative mechanism where candidates optimally reposition in response to changes in the council size. Admittedly, we cannot be perfectly sure there is no other channel through which the council size could influence cohesion. Nevertheless, the electoral rule disproportionality is a plausible channel since our empirical results are in line with our conceptual framework and our balance tests do not reveal other evident determinants of cohesion. We also show that there is no sorting across the thresholds, which is natural as the municipal population is not self-reported. Last but not least, the results are robust to further battery of robustness and validity checks. Here of a particular interest is our novel use of placebo cutoff tests to assess the appropriate level of clustering in the optimal bandwidth selection.

The paper is structured as follows: In Section 1.1 we present related literature and highlight our contribution. In Section 2 we propose a conceptual framework to guide our empirical analysis. In Section 3 we present our estimation strategy and the evidence. In Section 4, we conclude. All supporting material, further discussion of our data and empirical results (e.g., robustness), as well as a formal analysis of our conceptual framework appear in the Online Appendix published as Supplemental Material available on the journal's website.

## 1.1 Contribution

Our estimates on the effect of electoral institutions on parties' ideological cohesion constitute a novel finding. In a broader context, evidence regarding the causal effects of electoral institutions on any outcome remains relatively scarce (Shugart 2013). Typically, researchers have had to rely on cross-country or panel variation leaving room for potential confounding. In contrast, we leverage on plausibly as-good-as-random variation within a country by utilizing the change in council size as a proxy for electoral rule disproportionality and join scholars exploiting subnational variation to identify causal effects of various dimensions of electoral institutions (see e.g., Eggers 2015; Sanz 2017).

As Shugart (2013) and Buisseret and Prato (2017) argue, we still have a limited overall understanding of the mechanisms linking electoral institutions and the behavior of political actors. We contribute to this literature by proposing a potential mechanism through which electoral institutions could influence parties' candidate selection on ideology. Our theoretical framework builds on recent literature; modeling proportional representation as a system of power sharing, as compared to majoritarian winner-take-all elections (Herrera et al. 2014, 2016, 2019; Matakos et al. 2016) and extends knowledge on parties' cohesion across electoral institutions (Carey 2007; Buisseret and Prato 2022).

Closest to our setup, Matakos et al. (2016) propose an electoral competition model where two parties optimally choose their platforms on a unidimensional policy space (i.e., each party selects a unique point on the unit interval) under different levels of electoral rule disproportionality. Their main result suggests that polarization (i.e., the distance between parties' platforms) decreases as electoral disproportionality increases, since parties have incentives to increase their vote share by converging towards the median. In our setup instead, parties propose an interval of policies (i.e., a list) that takes into account parties' preferences over the overall composition of the elected body. In this setting, we introduce a personal vote dimension and characterize the optimal length of parties' lists as the electoral rule disproportionality varies.[1] While Matakos et al. (2016) provide cross-country evidence on the correlation between the electoral rule disproportionality and between-party polarization, we focus on a within country setting and obtain causal evidence on the effects of council size as a proxy of the electoral rule disproportionality on intraparty cohesion.

Overall, our work contributes towards a better understanding of politicians' characteristics in representative democracies.[2] While the elected officials' characteristics and policy positions are known to matter for policy outcomes in several environments (see e.g., Chattopadhyay and Duflo 2004; Lee et al. 2004; Jones and Olken 2005; Washington 2008; Besley et al. 2011) —including the context of our empirical exercise and via an intraparty channel (Hyytinen et al. 2018a), different institutions seem to elect politicians with dissimilar traits (Best et al. 2000; Meriläinen 2022). Variation in electoral institutions has been offered as one potential explanation for differences in political selection (Carey and Shugart 1995; Galasso and Nannicini 2011; Gagliarducci and Nannicini 2013; Beath et al. 2016; Galasso and Nannicini 2017), with the present paper being, to the best of our knowledge, the first to provide causal evidence on electoral institutions affecting political selection on ideology. This result, hence, complements a long literature highlighting electoral institutions as a key determinant of the economy and the political

[1]Our results would not be qualitatively distinct from those of Matakos et al. (2016) if we *entirely* removed the personal vote element. In the Appendix we provide extensions of our main setting guaranteeing robustness of our results when voters *also* care about the lists' mean ideology or compare the expected utility obtained by the two lists.

[2]The focus of our paper is on intraparty heterogeneity at the candidate level. Nevertheless, our main result transcends from the stage of candidate selection (and list composition) to the ideological composition of the elected council (see Table A6).

system in representative democracies.[3]

## 2 A conceptual framework

Electoral rules define the allocation of a number of seats available in an elected body (e.g., a council or parliament) to parties competing in an election. That is, electoral rules are defining a mapping from votes shares to seats shares. Generating a fully proportional mapping where seat shares and votes shares coincide is practically impossible due to distortions emerging when dividing a discrete number of seats. However, specific features of the electoral rule generate diverse distortions. A large literature focuses on such specific characteristics of electoral rules (Herron et al. 2018).[4] Among those, one is the size of the elected body, the focus of our empirical application. Imagine that only one candidate is elected. One vote more than the opponent is enough and the distortion in the mapping from votes to seats in maximized. Instead, when many candidates are elected the seat share allocation is expected to generate less distortions.

In our theoretical framework, we abstract from the exact source creating distortions in the vote to seat share mapping. We follow conventional terminology by referring to the electoral rule disproportionality *as the expected distortions generated by the rule in favor of the election winner* and discuss a channel through which electoral rule disproportionality could affect parties' intraparty cohesion. Our argument is based on standard assumptions of the electoral competition literature and, while presented informally here,

[3]The literature on electoral institutions is vast. For some representative examples among many others, we refer the reader to studies on the effect of electoral institutions on polarization (Cox 1990), turnout (Blais and Carty 1990; Herrera et al. 2014), campaign spending (Iaryczower and Mattozzi 2013), corruption, redistribution, public spending and the provision of public goods (Lizzeri and Persico 2001; Milesi-Ferretti et al. 2002; Persson et al. 2003, 2007). For more references see therein as well as Persson and Tabellini (2002, 2005); Lijphart (1995, 1999); Taagepera and Shugart (1989); Grofman (2008).

[4]Even in the class of proportional representation rules, the exact allocation method decisively affects the seat allocation. For instance, while the D'Hondt method present in our empirical setting, "tends to increase the advantage for the electoral lists which gain most votes to the detriment of those with fewer votes" (Kotanidis 2019, p. 1), other methods generate less disproportional results (e.g., Hare-Niemeyer and Sainte-Laguë methods, see Schuster et al. 2003). Additionally, further explicit distortions in favor of the election winner, aimed at strong governments, can be introduced through the presence of a minimum vote threshold to obtain representation, a bonus to the winner of the election, or others (Herron et al. 2018).

it is derived in full mathematical detail in the Appendix in Section A1. In the Appendix, we also detail modifications that would not affect the intuition behind the result presented (Section A1.1). However, let us clarify that our main result essentially depends on the following key premises: On the voters' side, our model requires that a personal vote dimension is present.[5] In our benchmark model, voters vote for the party that includes in its list the candidate closest to their ideal point. In the Appendix, we show that our results are robust when permitting voters to also care about the average ideology of a given list or introduce some voters that choose the list that offers them higher expected utility. On the parties' side instead, our model requires that, ceteris paribus, parties prefer to nominate candidates who share the parties' ideology, and parties have full power in nominating their party lists.[6]

**The electoral rule and disproportionality:** To fix ideas and provide meaningful comparative statics, our conceptual framework below presents the simplest case of a linear electoral rule (Taagepera 1986). This rule serves as a continuous approximation of the D'Hondt rule present in our empirical setting (Flis et al. 2020), and links with our empirical measures of realized distortions. In the Appendix, we present further details on the D'Hondt rule and obtain theoretical results for the latter as well as a larger family of electoral rules than the linear rule.

To formalize the electoral rule and its degree of disproportionality, consider a two-

[5] A long-standing literature on personal vote has found that many voters condition their electoral choices on individual candidates' characteristics, attributes and promises (proposals) in a variety of countries and institutional contexts (for a summary see seminal contributions such as King 1991; Carey and Shugart 1995; Cain et al. 2013). This, in turn, generates strong incentives for candidates to seek for and cultivate their "personal" vote. Such incentives are further intensified in open-list multi-member district PR systems, like in Finland (see e.g Carey and Shugart 1995). Moreover, in Finnish municipal councils there is no ruling coalition and the role of parties as legislative teams is relatively weak with several voters liberated to adopt a candidate-oriented perspective in deciding how to vote. For example, in the 2017 Finnish Municipal Election Survey a majority of the respondents stated that candidates were more important than parties in determining their vote (Borg 2018). Personal vote is also important in Finnish parliamentary elections with 42-51% of the voters voting based on the candidate instead of party in surveys covering five parliamentary elections up to 2011 (see e.g Borg 2012, Table 18.2).

[6] Party-centered process of candidate selection is highly relevant for the Finnish context (even at the local level). While it is possible to run as a candidate independently of parties if the aspiring candidate can collect 10 signatures from local eligible voters in support of the candidacy, this is very rare in practice: In 2012 municipal elections only 2.2% of candidates were independent.

party election where party $j$ obtains vote share $v_j \geq 0$ and party $-j$ obtains vote share $v_{-j} = 1 - v_j \geq 0$. The electoral rule in our model is mapping each party's vote share $v_j$ into a seat share $s_j(v_j) \geq 0$ with $s_j + s_{-j} = 1$. The *Linear rule* is formally defined as follows:

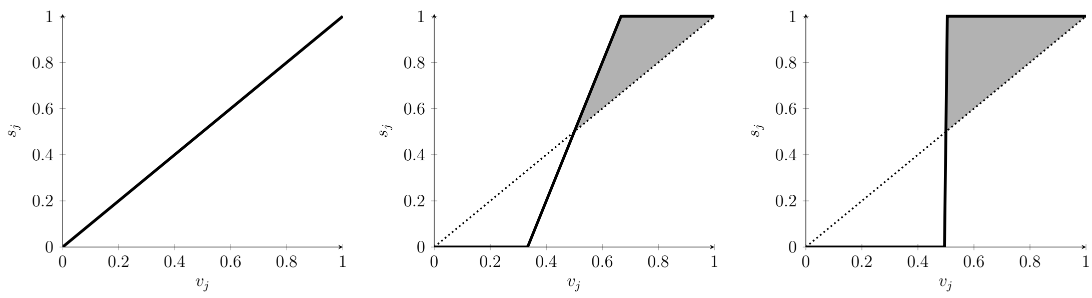
$$s_j(v_j) = \begin{cases} 0 & \text{if } v_j \leq \frac{n-1}{2n} \\ \frac{1-n}{2} + nv_j & \text{if } \frac{n-1}{2n} < v_j \leq \frac{1+n}{2n} \\ 1 & \text{if } v_j > \frac{1+n}{2n} \end{cases}$$

with $n \geq 1$ (see Figure 1).

We define the electoral rule disproportionality as the expected distortions generated by the rule in favor of the election winner, formally defined as follows:

$$D = \int_{0.5}^{1} (s(v_j) - v_j) dv_j$$

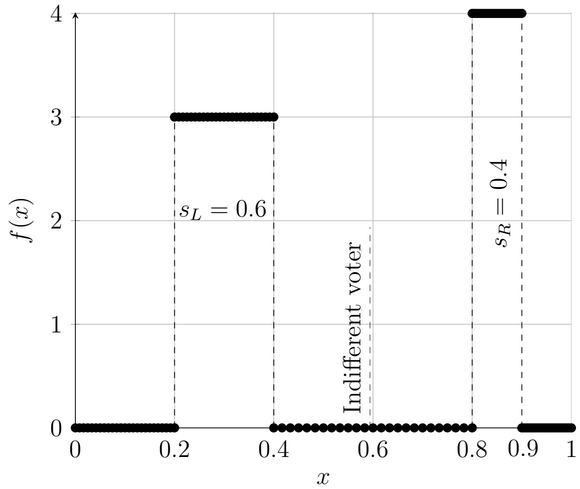
For the linear rule where $s(v_j) \geq v_j$, we have that $D = 0.125 - (0.125/n)$. This makes evident that the electoral rule disproportionality is increasing in parameter $n$. This is illustrated in Figure 1. When $n = 1$, parties' seat shares accurately represent parties' vote shares and the disproportionality is equal to zero. The extreme case of maximal disproportionality occurs in a winner-take-all election where $n \to \infty$.



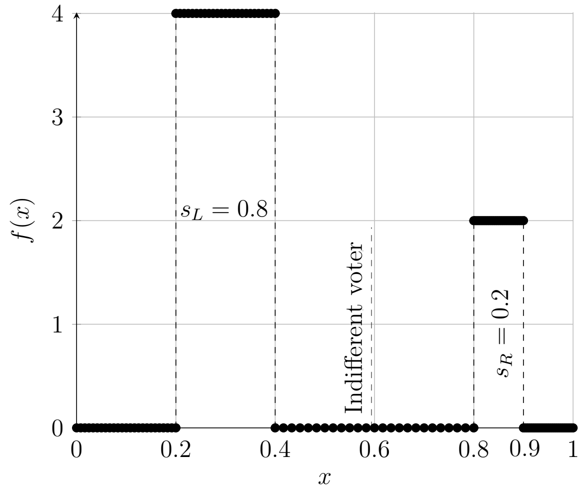

(a) $n = 1$ (b) $n = 3$ (c) $n \to \infty$

Figure 1: Seat share allocation given parties' vote shares according to the Linear rule where the electoral rule disproportionality $D$ (in gray) is increasing in $n$.

**The election, parties, and voters:** Consider an election in a unidimensional policy space where two parties ($j = L, R$) strategically choose the ideological heterogeneity of a continuum of candidates (*list*) competing in the election under the party's label. That is, candidates' ideologies are just numbers on the real line, and a list of candidates is given by a closed interval containing the party's ideal policy. For example, in Figure 2, the two parties propose lists $[0.2, 0.4]$ and $[0.8, 0.9]$ respectively.[7]




(a) Linear rule and $n = 1$.




(b) Linear rule and $n = 3$.

Figure 2: An example where parties propose $[\underline{x}_L, \overline{x}_L] = [0.2, 0.4]$ and $[\underline{x}_R, \overline{x}_R] = [0.8, 0.9]$. Parties' vote shares are $v_L = 0.6$ and $v_R = 0.4$. The electoral rule disproportionality determines parties' seat shares and hence the distribution of the represented ideologies in the elected body.

Voters' preferred policies are uniformly distributed on the same policy space. Voters are voting for the party that included the candidate closest to their preferred policy in its list. In the example of Figure 2, all voters to the left of the indifferent voter (located at 0.6) vote for party $L$, and all voters to the right of the indifferent voter vote for party $R$, implying vote shares $v_L = 0.6$ and $v_R = 0.4$.

**The elected body:** Parties' lists and seat shares $s_L$ and $s_R$ determine the distribution of ideologies in the elected body. The represented ideologies coincide with those of the nominated candidates, and each ideology has an equal weight within each party.[8] Take

[7]In our benchmark model parties propose non-overlapping intervals. One could relax this assumption without affecting the direction of our results. Consider party $L$ proposing interval $[\underline{x}_L, \overline{x}_L]$ and party $R$ proposing $[\underline{x}_R, \overline{x}_R]$ with $\overline{x}_L > \underline{x}_R$. Now all the voters in $[\underline{x}_R, \overline{x}_L]$ are indifferent between the two parties and, hence, randomize their vote. As $\overline{x}_L$ increases, then the vote share of party $L$ still increases but at a slower pace than in our model. Thus, the incentives to extend the list and to gain votes as the disproportionality increases are still present.

[8]Here we assume that the distribution of elected ideologies are uniformly distributed in the parliament

for example party $L$ in Figure 2a. Given that $n = 1$ this party is obtaining seat share $s_L = 0.6$, and each of the ideologies in $[0.2, 0.4]$ is represented in the elected body with an equal weight (i.e., 0.6/(0.4-0.2)). In contrast, in Figure 2b, since $n = 3$, party $L$ is obtaining $s_L = 0.8$, and each of the proposed ideologies $[0.2, 0.4]$ is represented in the elected body with a higher weight than before (i.e., 0.8/(0.4-0.2)).

**Parties' payoffs:** We assume that parties are policy motivated and their payoffs depend on the overall distribution of ideologies in the elected body. In specific, we consider that parties evaluate each ideology $t$ represented in the body using a quadratic disutility function, with their overall aim being to minimize the average disutility from the represented ideologies. Formally, party's $j \in \{L, R\}$ payoff, given the elected body, is

$$U_j([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \int_0^1 -(x_j - t)^2 f(t) dt$$

where $f(t)$ is the density of the represented ideologies as previously discussed and illustrated in Figure 2, and $x_j$ denotes the party's ideal policy.

## 2.1 Predictions

Parties propose a list of candidates running for a seat in the elected body (e.g., council or parliament) aiming at maximizing their payoff. So what determines parties' incentives to propose a more or less cohesive list of candidates?

A party has incentives to expand its list towards moderate grounds since the indifferent voter moves in its favor and the party attracts more votes. Any expansion towards moderate grounds in search of a higher vote share should be counterbalanced by the inclusion of more extreme candidates too, so that parties make sure that the elected candidates are not too moderate compared to the party's ideal policy. Hence, parties

despite the vote distribution being non uniform. Considering the continuous list as an approximation of a discrete list of equidistant candidates, this assumption requires that once the party elects more than two candidates, the two extreme and hence most voted candidates are definitely elected with the remaining seats distributed among the interior candidates with equal probability. For completeness of our argument, in Section A1 we show that our argument holds for other distributions.

benefit when increasing the ideological heterogeneity of their candidates by increasing their vote share and hence their seat share. However, proposing a heterogeneous list of candidates comes at a cost: the ideologies represented in the elected body may be too different from those parties advocate and, thus, negatively affect parties' payoffs, despite parties' optimally balancing their lists.

The above tradeoff and the incentives to propose a more or less cohesive list of candidates depend on the electoral rule disproportionality. Recall that, as the rule becomes more disproportional, the expected advantage for the election winner is increasing. That is, disproportional rules reward high vote shares more generously in terms of seat allocation than proportional rules. Therefore, parties incentives to become less ideologically cohesive are increasing in the electoral rule disproportionality since the expected benefits of doing so overweight the associated cost. The following comparative static result provides our main theoretical prediction formally presented in the Appendix.

**Prediction 1.** *Parties' lists of candidates become less ideologically cohesive as the electoral rule disproportionality $D$ increases.*

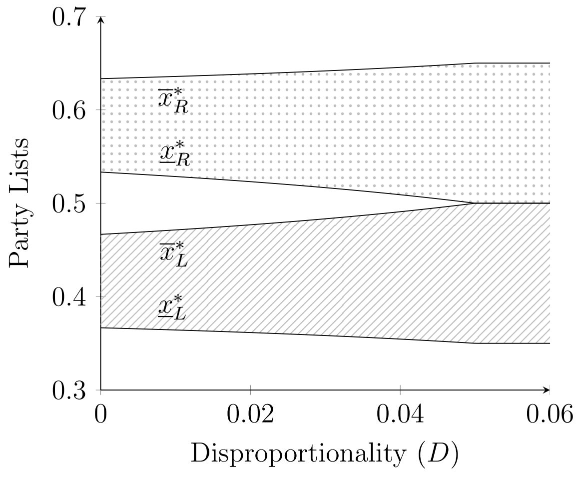



Figure 3: An example of equilibrium lists considering the Linear rule for different levels of disproportionality $D$ and parties' ideal policies $(x_L, x_R) = (0.4, 0.6)$. The two most extreme candidates in party $j$ are denoted by $\underline{x}_j^*$ and $\overline{x}_j^*$.

Figure 3 illustrates Prediction 1 plotting the ideology of the two most extreme candidates in each list, highlighting how parties' lists become less ideologically cohesive as the electoral rule disproportionality increases. Prediction 1 provides the main testable prediction of our theoretical model. In our empirical setting, we are using council size as a proxy of the electoral rule disproportionality and the above prediction implies that parties' lists of candidates are predicted to be more ideologically cohesive as the council size increases.[9]

# 3 Empirical evidence

We first describe the institutional details (Section 3.1), data of the empirical setup (Section 3.2), and then detail our identification strategy (Section 3.3). The same identification strategy is then used twice. First, we validate the use of council size as a proxy of the electoral rule disproportionality (Section 3.4). Second, we present our main results on the effect of the council size on parties' ideological cohesion (Section 3.5). Finally, we discuss alternative mechanisms that could potentially explain the effect of council size on cohesion instead of the electoral rule disproportionality (Sections 3.6 and 3.7) and provide further discussion on our methodology and robustness results (Section 3.8).

## 3.1 Institutional details

In our period of analysis, municipalities have a central role in the highly decentralized Finnish system. They spend more than 5,000 euros per capita annually and employ around 20% of the total workforce. Most of this expenditure is used to take care of statutory responsibilities, including social care, healthcare, and primary education. To cover these expenditures, Finnish municipalities are allowed to set and collect income taxes, property taxes and out-of-pocket payments from users of municipal services. In addition, municipalities receive a share of corporate taxes and fiscal grants from the central government. Therefore, municipalities wield a lot of power over public expenditures and revenues.

[9] In Appendix A1.1 we provide further theoretical details on the link between the electoral rule disproportionality and council size under the D'Hondt method present in our empirical setting.

Municipal councils are the main seat of power in the municipal decision-making system. No official ruling coalition government is formed after the elections, and councils decide by simple majority vote on an issue-by-issue basis. Relative to the parliamentary politics in Finland, Finnish local politics do not have very strong party discipline in place. Mayors are merely civil servants chosen by the council. The council also nominates a municipal board which has a preparatory role. Councilors are "leisure" politicians who receive small meeting fees while holding regular jobs. There is no evidence of large monetary gains from holding local office (Kotakorpi et al. 2017). Despite the small personal monetary stakes for the politicians, these are high-visibility elections that concern positions of power over important policies. Indeed, extensive existing causal evidence shows that individual councilors' characteristics matter for key policies in Finnish local governments (Hyytinen et al. 2018a; Harjunen et al. 2023; Meriläinen 2022).

Our conceptual framework was constructed with close parallels to Finnish municipal elections, where parties propose a list of candidates competing in an open list, and voters vote for a candidate who belongs to one of the lists. The number of candidates elected in each municipality (i.e., council size $k$) varies between 13 and 85 and is a deterministic step function of the municipal population.[10] In each municipality of council size $k$, parties propose an open list of up to $1.5 \times k$ candidates.[11] Candidates appear on the list in alphabetical order. Each voter votes for one candidate and cannot vote for a party without specifying a candidate. Candidates' personal votes are aggregated at the list level to determine lists' vote shares.[12] Lists' vote shares are mapped into lists' seats according to the D'Hondt allocation method. Seats are in turn allocated to the most voted candidates within a list. The D'Hondt method is the most common way of allocating seats

[10]The council sizes by municipal population are: Population of less than or equal to 2,000 (council size of 13, 15 or 17), 2,001-4,000 (21), 4,001-8,000 (27), 8,001-15,000 (35), 15,001-30,000 (43), 30,001-60,000 (51), 60,001-120,000 (59), 120,001-250,000 (67), 250,001-400,000 (75) and over 400,000 (85).

[11]Open lists are frequently encountered in national and subnational elections around the world using a PR system. For instance, at least forty countries, including many western European democracies use an open-list PR system when electing their single or lower chamber of their national parliament (Wall 2021).

[12]Parties can form pre-electoral coalitions and propose a joint list of candidates. The allocation of seats takes place at the coalition list level.

in proportional representation systems extensively used in western Europe (Kotanidis 2019). Two key elements of this rule are that: a) in expectation this rule favors the winner of the election (Gallagher 1991; Herron et al. 2018; Kotanidis 2019; Schuster et al. 2003), and b) the advantage to the winner is decreasing in the number of seats $k$ (Benoit 2000; Herron et al. 2018; Fiva and Hix 2019; Schuster et al. 2003; Kotanidis 2019).

## 3.2 Data sources

We combine data from several sources covering the Finnish municipal elections in 2008 and 2012. First, our key data on individual candidates' policy positions originate from the voting aid application (VAA) of the Finnish public broadcasting company, YLE. The YLE voting aid application is first open only to candidates who may reply to closed-ended questions focusing on current policy issues (see Section A4 in the Appendix for a detailed description). During the response period, each candidate has access only to their own replies, which can be modified during this time but not afterwards. Once the candidates' response period is over, the voting aid applications become publicly available. A voter can fill in the same questionnaire online and compare their replies to those of the candidates. The application also provides a list of candidates whose replies are closest to those of the voter. The open list makes Finland a fertile ground for the use of the voting aid application, as voters must find an individual candidate to vote for, so mere party-level information is not enough to guide the choice. Using the application is free of charge for both candidates and voters.[13] We have access to these data only for 2008 and 2012 elections.[14]

Given the importance of VAA in generating votes, candidates may have incentives for strategic responses. However, as the matching algorithms are trade secrets, they are not

[13]Finland is one of the first countries to introduce voting aid applications. They have gained popularity with surveys indicating that approximately 40% of the Finnish electorate used an application prior to the 2007 parliamentary election, with 15% of the users claiming that they had no favorite candidate and followed the application's recommendation (see Wagner and Ruusuvirta 2012 and references therein).

[14]Also the 2017 election data is available but the council size rule was no longer binding in those elections.

trivial to game. The strategic behavior by the candidates is further complicated by the fact that voter responses are not available, even afterwards. In addition, the responses of the candidates are fixed once the response period has ended so that candidates cannot react to other candidates' responses. Accordingly, Ilmarinen et al. (2022) show that the candidates responses are sincere rather than strategic by showing that candidates respond in the same way to a confidential survey as to the VAA.

Filling in the voting aid application questionnaire is not obligatory for the candidates. The median response rate by municipality in 2008 was 47.8% of the candidates and, on average, the candidates who did fill in a voting aid application questionnaire received 56.2% of the votes of the municipality.[15] The equivalent figures for 2012 were 47.2% of the candidates and 54.3% of the votes. Generally, the candidates who respond to the voting aid application are politically more successful, experienced, younger, more educated and more likely to be women (Table A9). However, as we show in Table 3, both the number of respondents and the number of all candidates (and thus, also the response rate) is balanced across the cutoffs used in the RDD and, hence, attrition is unlikely to pose any threat to our identification strategy. Nonetheless, we can still have the issue that different types of candidates respond across the cutoff. However, as detailed in Table A2, respondents' observed characteristics are also balanced across the cutoff.

Second, we use electoral data available from the Ministry of Justice with candidate-level information on candidates' age, gender, party affiliation, their election outcomes (number of votes and whether elected), and the possible incumbency status. These electoral data are linked to the data from Statistics Finland on candidates' education, occupation and socioeconomic status. Moreover, we match the candidate-level data with Statistics Finland's data on municipal characteristics. We have also collected information on pre-electoral coalitions of parties.

Using the electoral data, we construct our main disproportionality measures that we detail in Section 3.4. Similarly, using the YLE data, we construct the main outcome vari-

[15] Figure A12 reports the histograms of the share of respondents by municipality and by party.

ables on within-party cohesion that we detail in Section 3.5. All variables are summarized in Table A10.

### 3.3 Identification strategy and estimation

The deterministic council size rule allows for a sharp regression discontinuity design (RDD). The idea of our empirical strategy is to compare outcomes in municipalities just below and above the council size cut-off points.[16] The identifying assumption in such an RDD is that individuals cannot precisely manipulate the forcing variable (see e.g., Lee and Lemieux 2010). This is true in our case because municipalities do not self-report their population. The sufficient identification assumption is that the potential outcomes develop smoothly over the threshold.

We are interested mainly in two outcomes. First, we show that the council size has the expected effect on the distortions in the mapping of vote shares to seat shares. This validates that changes of the council size determined by the municipalities' population serves as a proxy of the electoral rule disproportionality. Second, as the main empirical contribution, we analyze whether there is an increase in intraparty cohesion at the threshold (that is, a discontinuous jump downwards in our within-party heterogeneity indices). Finally, we discuss other possible mechanisms that could explain the cohesion result.

To achieve this, we estimate regression models of these outcomes on a set of zero-one indicators for being above a cut-off point. We also include a flexible, but smooth, function of the population as control variables. The population variables should pick up the impact of all the determinants of within-party cohesion correlated with the population, apart from the council size. Hence, we will obtain a reliable estimate of the causal effect of the

[16] Regression discontinuity at population thresholds is a common approach to isolate causal effects. See for example Ferraz and Finan (2009); Egger and Koethenbuerger (2010); Gagliarducci et al. (2011); Fujiwara (2011); Pettersson-Lidbom (2012); Brollo et al. (2013); Gagliarducci and Nannicini (2013); Eggers (2015); Bordignon et al. (2016) among others. For a recent literature review and possible issues with the use of RDD at population thresholds see Eggers et al. (2018). We carefully address the concerns they raise. Similarly to us, Sanz (2017) and Lyytikäinen and Tukiainen (2019) use population thresholds to study political consequences of electoral systems.

council size on party cohesion, clean of confounding factors, that might otherwise bias our estimates.

As is standard in the literature, we use nonparametric local linear regressions as our main specification. We apply the bias correction and robust inference procedure by Calonico et al. (2014), which we implement using Calonico et al. (2016) rdrobust package in STATA. Based both on the Monte Carlo evidence by Calonico et al. (2014, 2018) and on an experimental benchmark by Hyytinen et al. (2018b), this approach performs best among the standard implementation options (that is, versus conventional local linear without the bias-correction and/or robust inference, and parametric polynomial specifications). We use the latest MSE-optimal bandwidth procedure proposed in Calonico et al. (2016) and apply triangular kernel.

We report the conventional local linear MSE-optimal coefficients, due to the method's optimal properties when it comes to point estimation. However, for statistical inference, and due to the superior coverage properties of the latter method, we report confidence intervals based on the bias-corrected coefficients and the associated robust inference by Calonico et al. (2014). This is somewhat non-standard reporting, as it implies that the reported 95% confidence interval is not centered precisely around the reported coefficient (but rather around the bias-corrected coefficient) but is, nonetheless, a well-motivated way to report. We report both classical and clustered inferences. The classical (non-clustered) inference has been standard in RDD for long as the typical optimal bandwidth selection methods have not been optimized for clustering. Due to the recent advances by Calonico et al. (2016), we can now also optimize the bandwidth selection while clustering. Note that, as opposed to the normal (non-RDD) case, clustering also changes the coefficients because the optimal bandwidths change.

One complication to our analysis is how to deal with multiple thresholds. One standard option is to calculate the forcing variable as a population distance to the nearest threshold and simply define a single group for being above a threshold. Given the limited amount of observations, we use this pooling option here. Cattaneo et al. (2016) show that, even if the pooling results in a loss of information, it produces meaningful (particularly weighted) treatment effect estimates. We can express this pooling approach as estimating regression functions of the form:

$$Y_{it} = \alpha + \delta \mathbb{1}_{z_{it}>0} + f(z_{it}) + \mathbb{1}_{z_{it}>0} f(z_{it}) + e_{it},$$

where $Y_{it}$ is the outcome of interest, $z_{it}$ is the forcing variable measuring the distance from the normalized population cutoffs for each observation $i$ in election $t$, $\mathbb{1}_{z_{it}>0}$ is an indicator function for being above a cutoff and $\delta$ is the coefficient of interest. If $f(z_{it})$ is approximately correctly specified within a bandwidth and there is no precise manipulation of the forcing variable (i.e., the density is smooth at the threshold), the covariates should evolve smoothly at the boundary and, thus, $\delta$ is be the causal estimate of interest.[17]

In all the analysis, we limit our sample to the municipalities with a population below 22,500, the mid-point between the fourth and fifth threshold. This focuses the analysis around the four smallest thresholds where the data is densest. Moreover, omitting larger cutoffs is theoretically motivated, as the relative changes in the council size are too small to induce a substantial treatment, that is, the changes to proportionality are very small. This also means that the treatment effect on cohesion should be very small at the large cutoffs (see Figure A4). By including such observations we arguably add more noise than information.[18] In addition, we omit the municipalities that underwent a municipal merger prior to the elections, as these have an impact on the council size and many other features of political competition. Overall, our sample contains 76% of all Finnish municipalities and 31% of the Finnish population.

Even if our pooling approach is standard in the literature, it is not entirely unprob-

[17]In the reported results, the bandwidth is optimized after pooling the data. However, the results are robust both to optimizing at each cutoff before pooling and to controlling for the cutoff fixed effects (not reported).

[18]However, the main results remain statistically significant in the non-clustered specifications even if we include the merged units and the larger municipalities but, as expected, the point estimates are closer to zero (Table A5). The results are also robust to limiting the sample further to 1-3 most densely populated cutoffs.

lematic. The main issue is that one could possibly end up comparing, for example, a municipality of a population of 1999 (just below) to a municipality of 8001 population (just above). This is clearly not a valid comparison for causal inference. Therefore, a further identifying assumption for pooling is that the share of identifying observations on both sides of each threshold is the same (which would happen in large samples due to local randomization). Thus, the McCrary (2008) density tests need to be reported separately for each threshold as opposed to the entire pooled sample. Additionally, we do not observe any jumps at any of the individual cut-offs or at the pooled one (see Figure A9 in the Appendix).

The standard identifying assumptions of our model imply that other possible determinants of intraparty cohesion should develop smoothly with respect to the population and, therefore, be captured by the $f$ function. This assumption is violated if there are other relevant factors that also depend on the same population rule. Eggers et al. (2018) have raised this concern especially in relation to the case of analyzing population thresholds since, in many countries, also municipal responsibilities, grants, and regulation as well as politicians' salaries depend on the very same thresholds. In such cases, there are several simultaneous exogenous treatments and RDD is able to identify only their joint effect. None of these concerns is present in the Finnish system. However, the council size in itself can have different electoral effects because candidates, parties, and voters may respond to it in various ways. To argue that the empirical mechanism is in line with our conceptual framework, we rely mainly on the covariate balance tests of the pre-treatment (before elections) variables (Section 3.6).

## 3.4 Council size and disproportionality

Note that the degree of disproportionality of an electoral rule —indicating the rule's tendency to favor parties obtaining high vote shares— is conceptually distinct from the precise advantage it confers upon those parties in a specific election.[19] However, given a large enough sample of electoral results, a more disproportional rule should be associated with a greater average advantage in favor of parties obtaining high vote shares. Indeed, in this section, we illustrate that the realized distortions favoring such parties in the mapping between vote shares and seat shares are, on average, more pronounced in smaller councils (see, also, Herron et al. 2018; Lijphart 1995). That is, we validate that council size is a good proxy of the electoral rule disproportionality.[20]

### Empirical measures of distortions

We use two established empirical measures of realized distortions in a certain election: i) the *Slope index*, directly linking to the Linear rule presented in our conceptual framework, and ii) a version of the *Gallagher index*. Both indices capture the distortions in the mapping of parties' vote shares to parties' seat shares and take larger values the higher is the distortion.[21] In total, we have 505 observations at the municipality-year level for which we compute the two indices as our main dependent variables.

The *Slope index* first proposed by Cox and Shugart (1991) is constructed as follows: For each municipality-year observation, we regress the difference $s_j - v_j$ on $v_j$. Then, we define the slope of the line obtained from this regression as the *Slope index*. The advantage of the slope index is that it captures the size of the distortions alongside whether they favor large or small parties. The Slope index is also linked to the way we have modeled the electoral rule disproportionality in our conceptual framework using the Linear rule.

Figure 4 illustrates how the slope of the regression line captures not only the size of

[19] As an illustration, consider elections using the linear rule with $n = 3$ as explained in our conceptual framework (see panel (b) in Figure 1) and two scenarios: one where the winner gets 55% of the vote and another where the winner gets 65%. With the same rule, the 55% vote results in a 65% seat share, while the 65% vote results in a 95% seat share. That is, while the disproportionality of the rule in both scenarios is the same, the realized distortions vary.

[20] According to our conceptual framework, the disproportionality of a D'Hondt rule that elects a $k$ member council is decreasing in $k$ (see the relevant subsection in Appendix A1.1).

[21] If parties have formed a pre-election coalition (roughly 15% of the lists in our sample) and, thus, run as a single joint list in the election, we define the list as a single party when calculating the distortions. This is to reflect the actual vote share to seat share mapping. When analyzing party cohesion, the RDD analysis on party cohesion could as well be conducted at the party level. While for consistency we report in the paper our cohesion analysis only at the coalition level, the results are similar at the party level.

the realized distortions but also the direction. On the left, we depict one municipality-year observation for which the distortions are small and the slope of the regression is relatively flat (0.011). In a pure PR rule, the slope would be zero. On the right, we depict another observation for which the slope is positive and relatively large (0.27), pointing to distortions favoring parties obtaining high vote shares. That is, the slope of the line used as our *Slope index* indicates whether distortions favor parties obtaining high vote shares (positive slope), parties obtaining low vote shares (negative slope) or the absence of distortions (flat line).

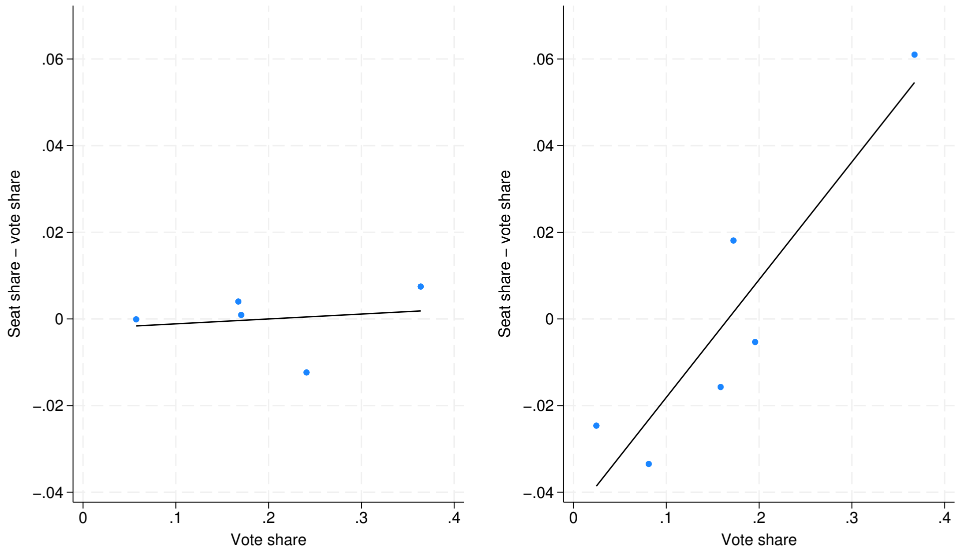



Figure 4: The *Slope index* as the slope of the regression of $s_j - v_j$ on $v_j$. The slope index takes the value of 0.0113 on the left (the Ilmajoki municipality in year 2012 with 5 competing lists and a council size of 35) and of 0.2712 on the right (the Utsjoki municipality in year 2008 with 6 competing lists and a council size of 21).

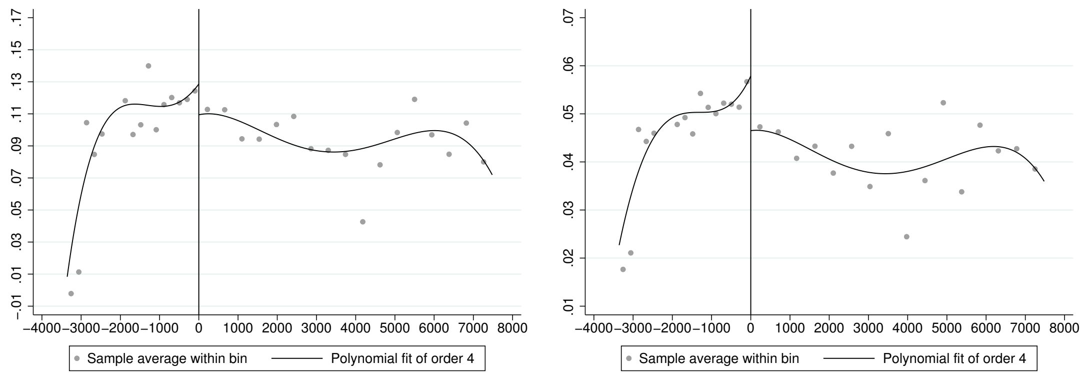
We employ the *Modified Gallagher index* as a second index to guarantee robustness of our results to alternative measures of realized distortions given that the early debate on the "best" way to capture the latter is still open (e.g., Lijphart 1995). The *Modified Gallagher index* in municipality $i$ in year $t$ is defined as:

$$MG_{it} = \left( \frac{1}{2} \sum_{j=1}^{p} \left( \frac{s_j}{\left(\sum_{j=1}^{p} s_j^2\right)^{0.5}} - \frac{v_j}{\left(\sum_{j=1}^{p} v_j^2\right)^{0.5}} \right)^2 \right)^{0.5}$$

where $j = 1, ..., p$ denotes the $p$ different parties running in municipality $i$ in year $t$. The difference in the summation term represents the distortions when a party $j$ that obtains vote share $v_j$ is allocated a seat share $s_j$. The value of the index is increasing in the level of distortions. In a pure PR system with no distortions, the index takes value zero since this difference is zero for all parties. Notice that, while the *Modified Gallagher index* (as most indices proposed in the literature) does well in representing the level of distortions in the vote-to-seat-share mapping, it remains silent on the direction of these distortions.[22]

**Distortions results**

We begin this RDD analysis with a graphical visualization of the jumps at the pooled cutoff. In Figure 5, we report the results for both indices of realized distortions. The results are very similar for both indices and both of them jump down at the cutoff.[23]




(a) Slope index. (b) Modified Gallagher index.

Figure 5: Pooled RDD. We use rdplot package in STATA. We report bins that mimic variance by using evenly-spaced method with spacings estimators. We use $4^{th}$ order polynomial for the fit.

In Table 1, we report the nonparametric RDD results on the effect of the council size on realized distortions. As for the main results, we report the conventional local linear

[22]The Modified Gallagher index is building upon the simpler Gallagher index (Gallagher 1991), possibly the most standard measure of distortions defined as: $G_{it} = (1/2 \times \sum_{j=1}^{p}(s_j - v_j)^2)^{1/2}$. Koppel and Diskin (2009) formalized the concerns by Taagepera and Grofman (2003) on the Galagher index showing that the modified version of the Gallagher index satisfies relevant properties that the Gallagher does not (e.g., Dalton's principles of transfers, scale invariance, orthogonality).

[23]We show that the results are similar when using a linear fit both for the whole sample and when limiting the sample within the MSE-optimal bandwidth in Figures A5 and A6.

MSE-optimal coefficients. For statistical inference, and because of its superior coverage properties, we report confidence intervals based on the bias-corrected coefficient and the associated robust inference by Calonico et al. (2014). We also report both the non-clustered results and those clustered at the municipality level. In line with the properties of the D'Hondt rule (see Appendix A1.1), the negative coefficients imply that distortions are smaller as the council size increases. All the coefficients have an expected negative sign. They are statistically significant at the 5% level for the Modified Gallagher index and (barely) insignificant for the Slope index. The magnitude of the effect of crossing the threshold translates into a decrease in the Slope index (Modified Gallagher index) by roughly 21% (26%) relative to the mean value of the outcome, or 40% (67%) of the outcome standard deviation.

Table 1: Realized distortions and council size, nonparametric RDD

| | **Slope index** | |
|---|---|---|
| Conventional local linear RD coefficient | -0.023 | -0.023 |
| 95% Confidence interval with bias-correction and robust inference | [-0.063 ; 0.007] | [-0.062 ; 0.007] |
| N within main bandwidth | 267 | 268 |
| MSE-optimal bandwidths (main/bias) | 807/1385 | 811/1409 |
| Clustered bandwidths and s.e.'s | No | Yes |
| Outcome mean (std.dev.) | 0.108 (0.057) | |
| | **Modified Gallagher index** | |
| Conventional local linear RD coefficient | -0.012 | -0.012 |
| 95% Confidence interval with bias-correction and robust inference | [-0.022 ; -0.003] | [-0.023 ; -0.002] |
| N within main bandwidth | 239 | 240 |
| MSE-optimal bandwidths (main/bias) | 714/1097 | 724/1090 |
| Clustered bandwidths and s.e.'s | No | Yes |
| Outcome mean (std.dev.) | 0.047 (0.018) | |

Notes: Unit of observation is municipality-year. Results are generated using rdrobust package in STATA (Calonico et al. 2016). We use local linear regression to construct the point estimator. We use local quadratic regression to construct the bias correction. We use triangular kernel.

While it is a well-known fact in political science that council size (or district magnitude) serves as a proxy for electoral rule disproportionality (see e.g., Carey and Shugart 1995; Carey and Hix 2011), and this association is well established at a theoretical and descriptive level, there has been, to our knowledge, a lack of causal evidence on this relationship. Thus, our paper makes an additional independent contribution through these results, even if they are not its primary focus.

## 3.5 Main results: Council size and intraparty heterogeneity

We now analyze how the council size influences intraparty cohesion. Guided by our theoretical framework, we are expecting that lists become less heterogeneous as the council size becomes larger (Prediction 1).

### Empirical measures of cohesion

Our main outcome variables are two indices of candidate heterogeneity, based on candidates' responses to questions present in the voting aid application. The *All questions index* is constructed using all available responses. The *Redistribution index* focuses only on a subset of the questions focusing on taxation and redistribution (see Section A4 in the Appendix for the subset). The benefit of the comprehensive index is that it avoids selecting on questions. However, its drawback is that perhaps some of the questions are less relevant for the voters and the interpretation of this index is more complex. Therefore, we also focus on the redistribution index that is policy relevant and matches a classic left-right dimension in political economy.

We measure ideological heterogeneity using Euclidean distances (scaled by the number of questions) for both indices.[24] That is, we first compute for each candidate the distance between their response, and their *local* party mean response, for each question, and take a square of that. Then we sum these squared distances over all the questions included in the index and take a root of the sums of those squares. Finally, we divide this sum by the number of questions each year to reduce residual variance due to the fact that the

[24]There are obviously many other ways one could calculate similar indices. We have the luxury of using this simple and transparent metric as our interest is only in the static relative position of a candidate in relation to their party. Moreover, we cannot use, for example, Mahalanobis or similar standardized distance measures at the within-party level because, by construct, they would force all the lists to be about equally cohesive.

number of questions differs by year.[25] Our scaling makes not only the mean, but also the entire distribution, more comparable across the years, which is useful in the RDD analysis for computing optimal bandwidth and avoiding the need to control for a year fixed effect. If the distance is zero for a candidate, their ideology coincides with the local party mean. The larger the distance is, the more this candidate deviates from the local mean. We present descriptive statistics of the indices in Table A10 and their histograms in Figure 6.

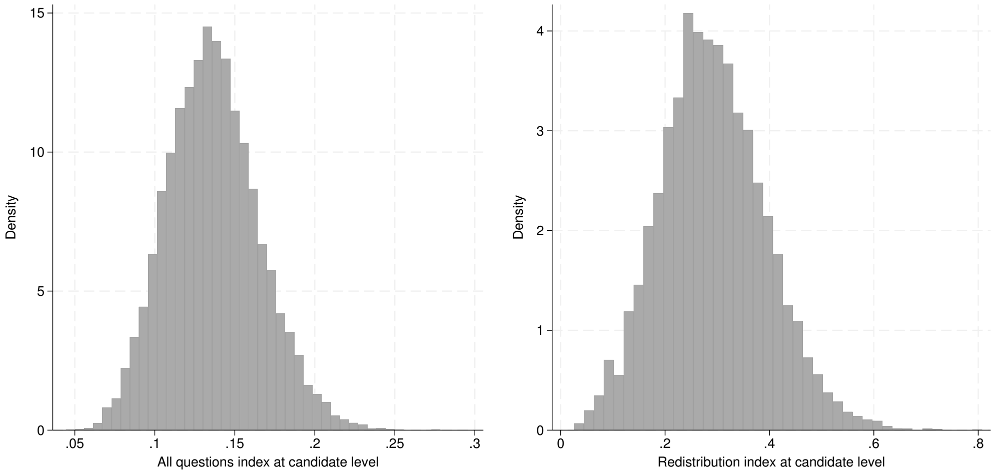



Figure 6: Histograms of the All questions and Redistribution indices at the candidate level.

For the analysis, we include only the parties with more than 5 candidates responding to the YLE voting aid application at the municipality-party-election level.[26] This leaves us with 14,999 candidate-election year, 1,184 party-election year and 475 municipality-election year observations.[27]

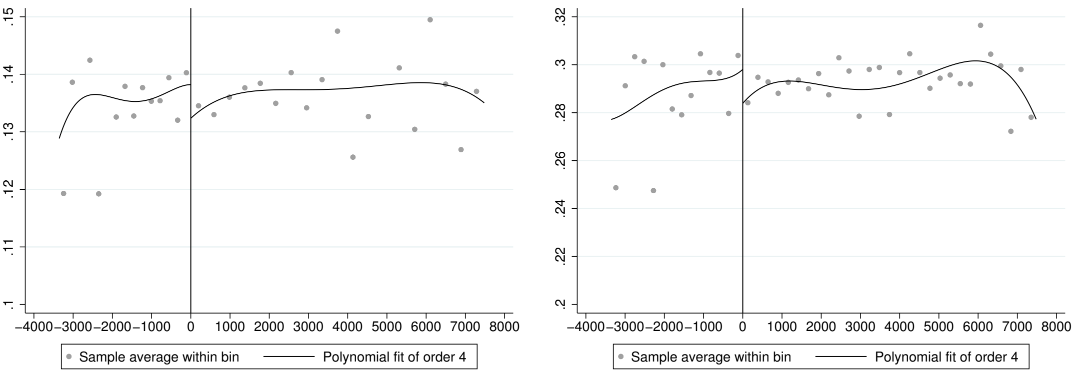
[25]We also implement a correction by multiplying the distance with $n/(n-1)$, where $n$ is the number of respondents. A correction similar to one used in computing sample variance.

[26]The probability that parties have more than 5 respondents does not change at the cutoff. The RDD effect (MSE optimal point estimate) on an indicator for "party has at least 5 candidates" is 0.02. This is small in magnitude and not statistically significant (robust and bias-corrected 95% CI is [-0.096 ; 0.138]).

[27]Note that, because we impose this minimum of five responses, we have 30 observations less at the municipal level than in the previous disproportionality analysis (Section 3.4).

## Cohesion results

We begin the main RDD analysis by graphical visualization of the jumps at the cutoff. In Figure 7, we report the results for the two heterogeneity indices using a pooled RDD and observations aggregated to municipality-year level. Both heterogeneity indices jump down at the threshold.[28] That is, parties become more ideologically cohesive as the council size increases.




(a) All questions index.

(b) Redistribution index.

Figure 7: Pooled RDD. We use rdplot package in STATA. We report bins that mimic variance by using evenly-spaced method with spacings estimators. We use $4^{th}$ order polynomial for the fit.

We present the nonparametric regression results in Table 2. Overall, the evidence is in line with our conceptual framework: The estimate is always negative indicating that party cohesion increases (that is, our dependent variable decreases) as the council size increases.

Panel A of Table 2 reports the analysis at the individual candidate level. Especially in this specification, clustering (optimal bandwidth selection and inference) at the municipality level is the most reliable approach as our treatment has no variation within the municipality-year level. To confirm that this does not give us excess power, we repeat the analysis at the municipality-party-year (panel B) and municipality-year (Panel C) level in Table 2. These aggregated outcomes are calculated as means over the individual

[28]We show that the results are similar when using a linear fit both for the whole sample and when limiting the sample within the MSE-optimal bandwidth in Figures A5 and A6.

Table 2: Policy positions and council size, nonparametric RDD

| | **All questions** | **All questions** | **Redistribution** | **Redistribution** |
|---|---|---|---|---|
| | **Panel A: Candidate-year level analysis** | | | |
| Conventional local linear RD coefficient | -0.012 | -0.009 | -0.023 | -0.021 |
| 95% Confidence interval with bias-correction and robust inference | [-0.018 ; -0.008] | [-0.019 ; -0.003] | [-0.043 ; -0.012] | [-0.043 ; -0.009] |
| N within main bandwidth | 2014 | 4256 | 3589 | 4357 |
| MSE-optimal bandwidths (main/bias) | 408/820 | 693/1143 | 588/1019 | 679/1118 |
| Clustered bandwidths and s.e.'s | No | Yes | No | Yes |
| Outcome mean (std.dev.) | 0.136 (0.029) | | 0.292 (0.101) | |
| | **Panel B: Municipality-party-year level analysis** | | | |
| Conventional local linear RD coefficient | -0.009 | -0.008 | -0.020 | -0.020 |
| 95% Confidence interval with bias-correction and robust inference | [-0.017 ; -0.004] | [-0.018 ; -0.001] | [-0.044 ; -0.004] | [-0.042 ; -0.006] |
| N within main bandwidth | 348 | 417 | 382 | 363 |
| MSE-optimal bandwidths (main/bias) | 617/1072 | 711/1152 | 663/1100 | 651/1100 |
| Clustered bandwidths and s.e.'s | No | Yes | No | Yes |
| Outcome mean (std.dev.) | 0.137 (0.015) | | 0.293 (0.042) | |
| | **Panel C: Municipality-year level analysis** | | | |
| Conventional local linear RD coefficient | -0.006 | -0.006 | -0.028 | -0.028 |
| 95% Confidence interval with bias-correction and robust inference | [-0.016 ; 0.003] | [-0.015 ; 0.002] | [-0.060 ; -0.007] | [-0.058 ; -0.009] |
| N within main bandwidth | 243 | 242 | 161 | 160 |
| MSE-optimal bandwidths (main/bias) | 813/1191 | 813/1206 | 558/956 | 550/958 |
| Clustered bandwidths and s.e.'s | No | Yes | No | Yes |
| Outcome mean (std.dev.) | 0.136 (0.013) | | 0.292 (0.032) | |

Notes: Results are generated using rdrobust package in STATA (Calonico et al. 2016). We use local linear regression to construct the point estimator. We use local quadratic regression to construct the bias correction. We use triangular kernel.

candidate distances aggregated to the respective level. The results are robust to these modifications.[29]

The effect of crossing the threshold in Table 2 translates into a decrease in the heterogeneity indices by roughly 7% (respectively 7%) relative to the mean value of the outcome. This effect also translates into a decrease of 20% (respectively 30%) of the outcome standard deviation for the redistribution index (respectively the all questions index).

## 3.6 Balance tests

We conduct balance tests on on a wide range of observables to check for other possible effects of moving across the population thresholds. In Table 3, we report the most important ones for our purposes and relegate further standard validity tests of RDD on municipality and candidate characteristics to the Appendix (Table A2).

Table 3: Balance tests

| | Number of parties | Effective numb. parties | Candidates per seat | Respondents | Candidates |
|---|---|---|---|---|---|
| Conventional local linear RD coefficient | 0.256 | 0.008 | -0.51 | 0.83 | 2.07 |
| 95% Confidence interval with bias-correction and robust inference | [-0.41 ; 1.15] | [-0.53 ; 0.58] | [-0.91 ; -0.21] | [-1.36 ; 2.84] | [-1.87 ; 5.98] |
| N within main bandwidth | 230 | 243 | 211 | 616 | 566 |
| MSE-optimal bandwidths (main/bias) | 778/1188 | 825/1187 | 705/1102 | 1067/1601 | 955/1367 |
| Clustered bandwidths and s.e.'s | Yes | Yes | Yes | Yes | Yes |
| Outcome mean (std.dev.) | 5.9 (1.4) | 3.5 (0.9) | 2.8 (0.7) | 12.7 (6.5) | 23.5 (11.3) |
| Unit of observation | Municipality-year | Municipality-year | Municipality-year | Party-year | Party-year |

Notes: For each outcome we use the smallest feasible level of aggregation at either municipal or local party level. Results are generated using rdrobust package in STATA (Calonico et al. 2016). We use local linear regression to construct the point estimator. We use local quadratic regression to construct the bias correction. We use triangular kernel.

First, it is important to stress that, across the cutoffs, the number of candidates and the number of parties (lists), either a simple count or an effective number of parties, are balanced (i.e., the effect is not statistically significant).[30] The balance in the number of parties is particularly relevant since, at a first sight, one could think that the changes in council size could also affect the number of parties. Note, however, that here we are

[29] In Table A1, we show that these results are robust to using a more flexible specification where we include cutoff specific fixed effects, allow different linear trends in the running variable around each cutoff, and optimize the bandwidths for each cutoffs separately before pooling the data.

[30] As is common in the literature, we compute the effective number of parties (lists) by the inverted Herfindahl index of vote shares of party lists.

focusing on the same PR setup across all the thresholds and, hence, the incentives known since the early work by Duverger (1954) may not be fully in place. Moreover, in the current context, most local parties are also well-known at the national level. Thus in practice, parties do not merge or split at the local level in Finland (unless the national party splits first). The main choice that affects the number of parties is whether the national party wants to enter or exit a given municipality. It is likely that the variations in the council size are not the parties' main concern when considering this major discrete choice. Entry and exit are likely rather driven, for example, by the existence of a large enough support base in the municipality. The balanced number of parties is also relevant because, as explained in Section A1.1, our conceptual framework is valuable in analyzing multi-party settings for an exogenous number of parties.

Additionally, the number of candidates is not significantly influenced at the cutoff. This highlights that parties and voters act in similar environments across the cutoffs. This is good news for our identification; yet one could expect jumps across the cutoffs since parties can nominate candidates up to 1.5 times of the council size.[31] The absence of jumps most likely implies that either new candidates are not sufficiently incentivized to run for office by the increase in the likelihood of getting elected due to the larger number of seats, and, hence, are not very strategic (see also Section 3.7); or that parties play an important role in curating their lists more efficiently at the face of increased candidate supply, thus facilitating more cohesion without any need to vary the number of candidates. Note that while the supply of candidates is certainly influenced by the municipal population, our research design controls for this by estimating the effect at the cutoff, with municipalities being very similar across the thresholds both in population size and other respects (Table A2).

Finally, we point out that the response rate of the candidates on the YLE application is balanced by reporting also the balance test for the number of respondents. Here, in

[31] However this does not occur and lists usually include fewer candidates than the maximum permitted. Only 3.4% of lists are full in our sample. Typically only larger parties in larger municipalities do fill the lists completely.

correspondence with our sample from the main analysis, we also omit all candidates from parties with less than 5 respondents. Given that using the application is voluntary, one could be concerned about a possible selection bias. The balanced number of respondents (and given the balanced number of all candidates, also balanced response rate), however, indicates that a possible selection bias resulting from the response rate is unlikely to be present in the RDD estimates. As the possible selection bias seems to be the same across the cutoffs, it is thus differenced out from the RDD estimates. However, it could still be that there are differences in how the candidates select into the survey across the cutoff. To address this, we show the balance of candidate characteristics in the Appendix (Table A2).

Of course, we face the standard caveat of balance tests that the results may be statistically imprecise. In our case, we cannot rule out small effects of the council size on these outcomes. To further evaluate our cohesion results, we show in the Appendix (Table A3) that the disproportionality and the cohesion indices move significantly at the same cutoffs, whereas the alternative variables less so.

### 3.7 Strategic candidates

Our posited mechanism to explain our empirical results requires that parties are strategic and respond to the changes in electoral incentives (council size), while candidates are sincere. Yet, one cannot exclude an alternative mechanism; one that entails strategic candidates optimally repositioning in response to the change in the institutional setting. As the number of candidates remains about constant across the threshold, the available seats per candidate varies (see Table 3). Therefore, one could argue that as the number of votes necessary to win a seat goes down, and thus, intraparty competition becomes less intense, candidates face weaker incentives to diversify and distance themselves from each other. In other words, it is theoretically possible that the observed effects arise from candidates' rather than parties' strategic responses to changes in council size. While from a policy perspective it may not be relevant which agents are driving the link between council size and cohesion, a candidate driven mechanism is different from the one our model proposes, and this should be acknowledged.

Here, we present empirical evidence showing that the candidates, at least in the Finnish context, do not seem to act strategically in how they position. That is, while the alternative candidate centered mechanism is theoretically possible, and may well be important in other institutional setups, it is less likely to be the one driving our empirical results. Table A4 provides evidence in this direction by showing that candidates do not change their policy positions reported in the VAA application when their electoral incentives change. To achieve this, we analyze within candidate changes between 2008 and 2012. We focus on those VAA questions that remain the same across the years when the candidates' institutional environment changes. First, we analyze changes in council size, which may happen due to changes in municipal population, municipal mergers or candidate mobility. Whichever the reason, all should affect candidates' incentives to respond strategically. Second, we analyze party switching. Out of 8550 candidates in total that re-run and respond to both survey years, 5.3% switch their party. While candidates overall change their responses somewhat over time (*Constant*), candidates do not respond to changes in their contextual environment.[32] Only one out of 12 estimates of interest in Table A4 is statistically significant and the magnitudes of the point estimates in relation to the constants are small. Hence, Table A4 does not support the presence of strategic repositioning in our context. This result is actually in line with past studies of Finnish local politics: Savolainen (2020) has documented with a causal design that candidates do not change their VAA policy responses if they get elected. Moreover, non-strategic VAA responses seem natural in a setting where candidates have low office-seeking motives (see Section 3.1) while their election does affect policy. As Hyytinen et al. 2018a; Meriläinen 2022 show, there are substantial policy effects related to (quasi-randomly) electing individual candidates with certain characteristics.

At any rate, the two approaches are not mutually exclusive or in contradiction. Indeed,

[32] We study absolute changes so that possible changes in the opposite directions do not cancel each other out in the estimation.

as long as parties are strategic actors, they can select candidates in anticipation of the candidates' strategic positioning at any given institutional setting. Thus, it is likely that in various settings both mechanisms could operate in tandem.

### 3.8 Robustness, validity and discussion

In the Appendix (Section A2), we report and discuss the standard validity and robustness checks in detail. The McCrary (2008) test for manipulation shows no evidence on municipalities manipulating their population count at any individual cutoff, nor in the pooled data (Figure A9). This makes perfect sense because, since population counts are not self-reported by the municipalities, there are no incentives to manipulate this information. No other policies or municipality responsibilities change at these cutoffs. We also report that the main results are robust across a fair range of bandwidths around the optimal ones (Figure A10).

We report the placebo cutoff analysis in Figure A11 in the Appendix (Section A2). This analysis is especially useful for understanding whether the applied RDD specification is appropriate (Hyytinen et al. 2018b). Moreover, it shows that we should trust the clustered results much more than the non-clustered ones because there is within-municipality correlation in the policy positions of the candidates. If the bandwidth calculation does not account for this clustering problem, the optimal bandwidths are too narrow in the sense that the results are derived using only a couple of clusters. This leads to the standard problem that, in small samples, any result is possible by chance even if the design is as-good-as random. The placebo cutoff test for the clustered specifications works as it is supposed to and the coefficients are zero at the placebo cutoffs.

Finally, in Tables A7 and A8 we report 2SLS analysis where we estimate the effect of the realized distortions on intraparty cohesion while using the RDD cutoffs as instrumental variables. All coefficients across all specifications have the expected sign in the second stage (and the first stage and the OLS), that is, as the distortions increase party heterogeneity increases. However, the second stage results are not statistically significant given the low power of 2SLS in general and our weak first stage for 2SLS purposes. Importantly, since party lists are formed before the actual electoral distortion is realized, actual distortions are de facto not a determinant of party list cohesion. Thus, a 2SLS model (i.e., the council size affects the distortions at a given election, and these distortions affect party list cohesion) is arguably not the most suitable estimation approach in our setting.

# 4 Conclusion

Our work provides new insights on how electoral institutions alter parties' incentives when recruiting their political personnel and contributes to the literature on candidate selection and nomination (e.g., Besley et al. 2017; Dal Bò et al. 2017; Buisseret et al. 2022). As Dal Bò and Finan (2018) note, while parties appear to play a key role in political selection, the study of this intraparty dimension is still under-explored. In particular, the relationship between electoral institutions and intraparty ideological cohesion has been largely "black-boxed" in the literature; our study is one of the first to propose such link.

A better understanding of intraparty cohesion arguably carries non-trivial implications for several other closely related issues such as policy implementation. For instance, evidence shows that EU parliament legislation is sensitive to the ideological cohesion of the agenda setting party (Blumenau 2016). Similarly, the ability of national governments to implement important policies decreases when the governing party lacks ideological cohesion.[33] After all, even the probability of a party entering the government depends among others on its intraparty cohesion (see Tsebelis 2002 for theoretical arguments and Bäck 2008 for empirical evidence).

Moreover, intraparty cohesion can impact legislators' behavior in roll-call votes, serving as one of the two key elements, along with party whips, to ensure party unity within

[33] For indicative evidence refer to the cases of Italy regarding welfare state reforms (Ceron et al. 2019) and Germany during the 1980s and the low internal cohesion of the Christian democrats (Zohlnhöfer 2003).

legislatures. When intraparty cohesion is very high, parties vote cohesively as a block even in the absence of a party whip. Conversely, with low intraparty cohesion, no matter how strict the party whip may be, maintaining party unity becomes impossible. For intermediate levels, intraparty cohesion and party whip may act as substitutes (Bowler et al. 1999; Ceron 2015). Party unity in legislative voting in turn is relevant since not only by definition affects government stability but also affects government spending and resource allocation across districts with parties rewarding loyal voting behavior (Curto-Grau and Zudenkova 2018). Finally, our argument positing that proportional rules create incentives for parties to field ideologically homogeneous lists may provide fresh insights into the reasons behind the high levels of party unity observed in proportional representation systems (Cox et al. 2019).

## Data availability

Code replicating the tables and figures in this article can be found in Tukiainen et al. (2024) in the Harvard Dataverse, https://doi.org/10.7910/DVN/M1XHUT.

# References

Bäck, H. (2008). Intra-party politics and coalition formation: Evidence from Swedish local government. *Party politics 14*(1), 71–89.

Beath, A., F. Christia, G. Egorov, and R. Enikolopov (2016, 04). Electoral rules and political selection: Theory and evidence from a field experiment in Afghanistan. *Review of Economic Studies 83*(3), 932–968.

Benoit, K. (2000). Which electoral formula is the most proportional? A new look with new evidence. *Political Analysis 8*(4), 381–388.

Besley, T., J. G. Montalvo, and M. Reynal-Querol (2011). Do educated leaders matter? *Economic Journal 121*(554), F205–227.

Besley, T. J., O. Folke, T. Persson, and J. Rickne (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American Economic Review 107*(8), 2204–2242.

Best, H., M. Cotta, et al. (2000). *Parliamentary representatives in Europe, 1848-2000: Legislative recruitment and careers in eleven European countries*. Oxford University Press.

Blais, A. and R. Carty (1990). Does proportional representation foster voter turnout? *European Journal of Political Research 18*(2), 167–181.

Blumenau, J. (2016). Intra-party politics, cohesion, and legislative gridlock. *Manuscript*.

Bordignon, M., T. Nannicini, and G. Tabellini (2016). Moderating political extremism: Single round vs runoff elections under plurality rule. *American Economic Review 106*, 2349–2370.

Borg, S. (2012). Muutosvaalit 2011, omso 16/2012. *Helsinki: Oikeusministeriö*.

Borg, S. (2018). Finnish municipal election study 2017 [dataset]. Version 1.0 (2018-11-06). *Finnish Social Science Data Archive [distributor] http://urn.fi/urn:nbn:fi:fsd:T-FSD3251.*

Bowler, S., D. M. Farrell, and R. S. Katz (1999). Party cohesion, party discipline, and parliaments. In *Party discipline and parliamentary government*, pp. 3–22. Ohio State University Press.

Brollo, F., T. Nannicini, R. Perotti, and G. Tabellini (2013). The political resource curse. *American Economic Review 103*(5), 1759–1796.

Buisseret, P., O. Folke, C. Prato, and J. Rickne (2022). Party nomination strategies in list proportional representation systems. *American Journal of Political Science 66*(3), 714–729.

Buisseret, P. and C. Prato (2022). Competing principals? Legislative representation in list proportional representation systems. *American Journal of Political Science 66*(1), 156–170.

Buisseret, P. E. and C. Prato (2017). Electoral accountability in multi-member districts. *Manuscript*.

Cain, B., J. Ferejohn, and M. Fiorina (2013). *The Personal Vote.* Cambridge (MA): Harvard University Press.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association 113*(522), 767–779.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2016). rdrobust: Software for regression discontinuity designs. *Stata Journal 17*(2), 372–404.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Carey, J. M. (2007). Competing principals, political institutions, and party unity in legislative voting. *American Journal of Political Science 51*(1), 92–107.

Carey, J. M. and S. Hix (2011). The electoral sweet spot: Low-magnitude proportional electoral systems. *American Journal of Political Science 55*(2), 383–397.

Carey, J. M. and M. S. Shugart (1995). Incentives to cultivate a personal vote: A rank ordering of electoral formulas. *Electoral Studies 14*(4), 417–439.

Cattaneo, M. D., L. Keele, R. Titiunik, and G. Vazquez-Bare (2016). Interpreting regression discontinuity designs with multiple cutoffs. *Journal of Politics 78*(4), 1229–1248.

Ceron, A. (2015). Brave rebels stay home: Assessing the effect of intra-party ideological heterogeneity and party whip on roll-call votes. *Party Politics 21*(2), 246–258.

Ceron, A., L. Curini, and F. Negri (2019). Intra-party politics and interest groups: Missing links in explaining government effectiveness. *Public Choice 180*(3-4), 407–427.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica 72*(5), 1409–1443.

Cox, G. W. (1990). Centripetal and centrifugal incentives in electoral systems. *American Journal of Political Science 34*(4), 903–935.

Cox, G. W., J. H. Fiva, and D. M. Smith (2019). Parties, legislators, and the origins of proportional representation. *Comparative Political Studies 52*, 102–133.

Cox, G. W. and M. S. Shugart (1991). Comment on Gallagher's 'proportionality, disproportionality and electoral systems'. *Electoral Studies 10*(4), 348–352.

Curto-Grau, M. and G. Zudenkova (2018). Party discipline and government spending: Theory and evidence. *Journal of Public Economics 164*, 139–152.

Dal Bò, E. and F. Finan (2018). Progress and perspectives in the study of political selection. *Annual Review of Economics 10*(1), 541–575.

Dal Bò, E., F. Finan, O. Folke, T. Persson, and J. Rickne (2017). Who becomes a politician? *Quarterly Journal of Economics 132*(4), 1877–1914.

Duverger, M. (1954). *Political parties: Their organization and activity in the modern state*. Methuen.

Egger, P. and M. Koethenbuerger (2010). Government spending and legislative organization: Quasi-experimental evidence from Germany. *American Economic Journal: Applied Economics 2*(4), 200–212.

Eggers, A. C. (2015). Proportionality and turnout: Evidence from French municipalities. *Comparative Political Studies 48*(2), 135–167.

Eggers, A. C., R. Freier, V. Grembi, and T. Nannicini (2018). Regression discontinuity designs based on population thresholds: Pitfalls and solutions. *American Journal of Political Science 62*(‘), 210–229.

Ferraz, C. and F. Finan (2009). Motivating politicians: The impacts of monetary incentives on quality and performance. *NBER Working Paper No. 14906*.

Fiva, J. H. and S. Hix (2019). Electoral reform and strategic coordination. *British Journal of Political Science*, 1–10.

Flis, J., W. Słomczyński, and D. Stolicki (2020). Pot and ladle: A formula for estimating the distribution of seats under the Jefferson–D'Hondt method. *Public Choice 182*(1-2), 201–227.

Fujiwara, T. (2011). A regression discontinuity test of strategic voting and Duverger's law. *Quarterly Journal of Political Science 6*(3-4), 197–233.

Gagliarducci, S. and T. Nannicini (2013). Do better paid politicians perform better? Disentangling incentives from selection. *Journal of the European Economic Association 11*(2), 369–398.

Gagliarducci, S., T. Nannicini, and P. Naticchionia (2011). Electoral rules and politicians' behavior: A micro test. *American Economic Journal: Economic Policy 3*(3), 144–174.

Galasso, V. and T. Nannicini (2011). Competing on good politicians. *American Political Science Review 105*(1), 79–99.

Galasso, V. and T. Nannicini (2017). Political selection under alternative electoral rules. *Public Choice 171*(3-4), 257–281.

Gallagher, M. (1991). Proportionality, disproportionality and electoral systems. *Electoral Studies 10*(1), 33–51.

Grofman, B. (2008). The impact of electoral laws on political parties. In D. A. Wittman and B. R. Weingast (Eds.), *The Impact of Electoral Laws on Political Parties*. Oxford University Press.

Harjunen, O., T. Saarimaa, and J. Tukiainen (2023). Love thy (elected) neighbor? Residential segregation, political representation and local public goods. *Journal of Politics 85*(3).

Herrera, H., A. Llorente-Saguer, and J. C. McMurray (2019). The marginal voter's curse. *Economic Journal 129*(624), 3137–3153.

Herrera, H., M. Morelli, and S. Nunnari (2016). Turnout across democracies. *American Journal of Political Science 60*(03), 607–624.

Herrera, H., M. Morelli, and T. R. Palfrey (2014). Turnout and power sharing. *Economic Journal 124*(574), F131–F162.

Herron, E. S., R. J. Pekkanen, and M. S. Shugart (2018). *The Oxford handbook of electoral systems*. Oxford University Press.

Hyytinen, A., J. Meriläinen, T. Saarimaa, O. Toivanen, and J. Tukiainen (2018a). Public employees as politicians: Evidence from close elections. *American Political Science Review 112*(1), 68–81.

Hyytinen, A., J. Meriläinen, T. Saarimaa, O. Toivanen, and J. Tukiainen (2018b). When does regression discontinuity design work? Evidence from random election outcomes. *Quantitative Economics 9*, 1019–1051.

Iaryczower, M. and A. Mattozzi (2013). On the nature of competition in alternative electoral systems. *Journal of Politics 71*(3), 829–853.

Ilmarinen, V.-J., V. Isotalo, J.-E. Lönnqvist, and Åsa von Schoultz (2022). Do politicians' answers to voting advice applications reflect their sincere beliefs? Comparing publicly and confidentially stated ideological positions in a candidate-centred electoral context. *Electoral Studies 79*, 102504.

Jones, B. F. and B. A. Olken (2005). Do leaders matter? National leadership and growth since World War II. *Quarterly Journal of Economics 120*(3), 835–864.

King, G. (1991). Constituency service and incumbency advantage. *British Journal of Political Science 21*(1), 119–128.

Kirchheimer, O. (1966). The transformation of the Western European party systems. In J. LaPalombara and M. Weiner (Eds.), *Political Parties and Political Development*. Princeton: Princeton University Press.

Koppel, M. and A. Diskin (2009). Measuring disproportionality, volatility and malapportionment: Axiomatization and solutions. *Social Choice and Welfare 33*(2), 281–286.

Kotakorpi, K., P. Poutvaara, and M. Terviö (2017). Returns to office in national and local politics: A bootstrap method and evidence from Finland. *Journal of Law, Economics, and Organization 33*(3), 413–442.

Kotanidis, S. (2019). Understanding the D'Hondt method. allocation of parliamentary seats and leadership positions. *European Parliamentary Research Service Briefing PE 637.966*.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355.

Lee, D. S., E. Moretti, and M. J. Butler (2004). Do voters affect or elect policies? Evidence from the US house. *Quarterly Journal of Economics 119*(3), 807–859.

Lijphart, A. (1995). *Electoral Systems and Party Systems*. Oxford University Press, USA.

Lijphart, A. (1999). *Patterns of democracy: Government forms and performance in thirty-six democracies*. New Haven: Yale University Press.

Lizzeri, A. and N. Persico (2001). The provision of public goods under alternative electoral incentives. *American Economic Review 91*, 225–239.

Lyytikäinen, T. and J. Tukiainen (2019). Are voters rational? *European Journal of Political Economy 59*, 230–242.

Matakos, K., O. Troumpounis, and D. Xefteris (2016). Electoral rule disproportionality and platform polarization. *American Journal of Political Science 60*(4), 1026–1043.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Meriläinen, J. (2022). Political selection and economic policy. *The Economic Journal 132*(648), 3020–3046.

Milesi-Ferretti, G. M., R. Perotti, and M. Rostagno (2002). Electoral systems and public spending. *Quarterly Journal of Economics*, 609–657.

Persson, T., G. Roland, and G. Tabellini (2007). Electoral rules and government spending in parliamentary democracies. *Quarterly Journal of Political Science 2*(2), 155–188.

Persson, T., G. Tabellini, and F. Trebbi (2003). Electoral rules and corruption. *Journal of the European Economic Association 1*(4), 958–989.

Persson, T. and G. E. Tabellini (2002). *Political economics: Explaining economic policy*. MIT press.

Persson, T. and G. E. Tabellini (2005). *The economic effects of constitutions*. MIT press.

Pettersson-Lidbom, P. (2012). Does the size of the legislature affect the size of government? Evidence from two natural experiments. *Journal of Public Economics 96*(3), 269–278.

Rae, D. (1967). *The political consequences of electoral laws*. Yale University Press.

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

Savolainen, R. (2020). Does political experience influence policy positions? *Manuscript*.

Schuster, K., F. Pukelsheim, M. Drton, and N. R. Draper (2003). Seat biases of apportionment methods for proportional representation. *Electoral Studies 22*(4), 651–676.

Shugart, M. S. (2013). Between science and engineering: Reflections on the APSA presidential task force on political science, electoral rules, and democratic governance: Why ballot structure matters. *Perspectives on Politics 11*(3), 818–820.

Taagepera, R. (1986). Reformulating the cube law for proportional representation elections. *American Political Science Review 80*(2), 489–504.

Taagepera, R. and B. Grofman (2003). Mapping the indices of seats–votes disproportionality and inter-election volatility. *Party Politics 9*(6), 659–677.

Taagepera, R. and M. S. Shugart (1989). *Seats and votes: The effects and determinants of electoral systems*. Yale University Press New Haven.

Tsebelis, G. (2002). *Veto Players: How Political Institutions Work*. Princeton, NJ: Princeton University Press.

Tukiainen, J., K. Matakos, R. Savolainen, O. Troumpounis, and D. Xefteris (2024). *Replication Data for: 'Electoral Institutions and Intraparty Cohesion'*. Harvard Dataverse, https://doi.org/10.7910/DVN/M1XHUT.

Wagner, M. and O. Ruusuvirta (2012). Matching voters to parties: Voting advice applications and models of party choice. *Acta Politica 47*(4), 400–422.

Wall, A. (2021). *Open List Proportional Representation: The Good, the Bad and the Ugly*. International Institute for Democracy and Electoral Assistance.

Washington, E. L. (2008). Female socialization: How daughters affect their legislator fathers. *American Economic Review 98*(1), 311–32.

Zohlnhöfer, R. (2003). Partisan politics, party competition and veto players: German economic policy in the Kohl era. *Journal of Public Policy 23*(2), 123–156.

# Online Appendix for: "Electoral Institutions and Intraparty Cohesion"

Konstantinos Matakos[*] Riikka Savolainen[†] Orestis Troumpounis[‡]
Janne Tukiainen[§] Dimitrios Xefteris[¶]

May 6, 2024

[*]King's College London, Strand campus, London, WC2R 2LS, UK

[†]Swansea University, Bay Campus, Fabian Way, Swansea SA1 8EN, United Kingdom.

[‡]University of Padova, via del Santo 33, 35123 Padova, Italy and Lancaster University, Bailrigg, Lancaster, LA1 4YX, United Kingdom.

[§]*Corresponding author:* Turku School of Economics, Rehtorinpellonkatu 3, FI-20014 University of Turku; and VATT Institute for Economic Research, Arkadiankatu 7, Helsinki FI-00101. Email: janne.tukiainen@utu.fi

[¶]University of Cyprus, PO Box 20537, 1678 Nicosia, Cyprus.

# Appendix

## A1 Formal analysis

Our conceptual framework presented in Section 1 is based on a formal model of electoral competition in which two parties ($j = L, R$) strategically choose the ideological heterogeneity of a continuum of candidates (*list*) competing in the election under the party's label. Given parties' lists, voters vote for one of the two parties and the electoral outcome is determined. The electoral rule in place is then determining the mapping of the electoral outcome to the elected body and parties' payoffs are realized.

We present the simplest version of our model that presents a set of interesting results on intraparty cohesion. In Section A1.1 we discuss our main assumptions and several ways we could relax some of those without changing the qualitative features of our equilibrium.

**Preliminaries**: The policy space is continuous, unidimensional, and represented by the interval $X = [0, 1]$. Parties have ideal policies, $x_L$ and $x_R$, that are symmetric around 0.5 (i.e., $x_L + x_R = 1$) with $x_L \in [1/3, 1/2]$ and $x_R \in [1/2, 2/3]$. Each party $j$ strategically chooses an interval $[\underline{x}_j, \overline{x}_j]$ where its candidates belong to maximize the party's payoff, with $0 \leq \underline{x}_L < \overline{x}_L \leq 1/2$ and $1/2 \leq \underline{x}_R < \overline{x}_R \leq 1$.[1]

The ideal policies of a unit mass of voters are uniformly distributed on the policy space, with $x_i$ denoting the ideal policy of individual $i$. In our benchmark model, we assume that voters vote for the party that included in its list the ideologically closest candidate to them. Since voters' behavior is parametric, we focus on symmetric Nash Equilibria in pure strategies in the list selection stage. A symmetric equilibrium is essentially a pair of intervals $[\underline{x}^*_L, \overline{x}^*_L]$ and $[\underline{x}^*_R, \overline{x}^*_R]$ such that $\overline{x}^*_L + \underline{x}^*_R = 1$, $\underline{x}^*_L + \overline{x}^*_R = 1$ and none of the two parties has an incentive to propose a different list of candidates.

[1]Assuming that parties propose a non-degenerate interval is purely for expositional reasons. As we actually show in the proof of our main result, in the equilibrium, parties would never propose a point on the strategy space. Also, the restriction on parties' ideal policies just guarantees that parties do not hit the extremes of the policy space and hence simplifies the analysis since it permits us to avoid focusing on the less interesting corner solutions.

**The electoral rule, vote and seat shares:** The electoral rule $s_j(v_j) : [0,1] \to [0,1]$ in our model maps each party's vote share $v_j$ into a seat share $s_j$ in the elected body. Let us first describe how $v_j$ is determined given the proposed intervals $[\underline{x}_L, \overline{x}_L]$ and $[\underline{x}_R, \overline{x}_R]$. The indifferent voter is located at $\frac{\overline{x}_L+\underline{x}_R}{2}$. All voters to the left (right) of the indifferent voter identify the closest candidate to their ideal policy in the list proposed by the leftist (rightist) party. Given the uniform distribution of voters, parties' vote shares are:

$$v_L = \frac{\overline{x}_L + \underline{x}_R}{2} \text{ and } v_R = 1 - \frac{\overline{x}_L + \underline{x}_R}{2}.$$

We first capture the electoral system by the *linear rule* (e.g., Taagepera 1986), a rule also used as the continuous approximation of the D'Hondt rule present in our empirical setting (Flis et al. 2020). In Section A1.1, we extend our analysis to other rules. Figure 1 in the main text illustrates distortions in favor of the winner according to the linear rule. Formally,

$$s_j(v_j) = \begin{cases} 0 & \text{if } v_j \leq \frac{n-1}{2n} \\ \frac{1-n}{2} + nv_j & \text{if } \frac{n-1}{2n} < v_j \leq \frac{1+n}{2n} \\ 1 & \text{if } v_j > \frac{1+n}{2n} \end{cases}$$

where $n \geq 1$. One can interpret this this rule as a linear mapping from vote shares to seat shares where the loser of the election is required to reach a threshold $(n-1)/2n$ to obtain representation. In Figure 1, we present two extreme cases: The absence of a vote threshold leads to a pure PR system ($n = 1$) while a winner-take-all election ($n \to \infty$) has an almost 50% threshold. An intermediate case with $n = 3$ is also presented. Clearly, the larger is $n$, the steeper is the slope of the rule, implying a larger advantage for the election winner.

**Disproportionality**: We consider the most proportional mapping to be the one in which vote shares accurately map to seat shares (i.e., the 45 degrees line, $n = 1$ for the Linear rule) while the "winner takes it all" is the most disproportional mapping (i.e., a two-step function, $n \to \infty$ for the Linear rule). We define the *electoral rule* *disproportionality as the expected distortions generated by the rule in favor of the election winner if all electoral results were equally likely*, formally defined as follows:

$$D = \int_{0.5}^{1} (s(v_j) - v_j) dv_j$$

For the linear rule where $s(v_j) \geq v_j$, we have that $D = 0.125 - (0.125/n)$, making evident that the electoral rule disproportionality is increasing in parameter $n$.

**Elected Body:** Given parties seat shares $s_L$ and $s_R$, we can determine the distribution of ideologies of the members of the elected body. Formally, the distribution of ideologies has support on $[\underline{x}_L, \overline{x}_L] \cup [\underline{x}_R, \overline{x}_R]$ (the list chosen by the parties), and ideologies are uniformly distributed within each party, with the density given by the following function (for an example see Figure 2:

$$f(x) = \begin{cases} 0 & \text{if } x < \underline{x}_L \\ s_L/(\overline{x}_L - \underline{x}_L) & \text{if } \underline{x}_L \leq x \leq \overline{x}_L \\ 0 & \text{if } \overline{x}_L < x < \underline{x}_R \\ s_R/(\overline{x}_R - \underline{x}_R) & \text{if } \underline{x}_R \leq x \leq \overline{x}_R \\ 0 & \text{if } x > \overline{x}_R \end{cases}$$

**Parties' Payoffs:** We assume that parties' payoffs depend on the distribution of ideologies in the elected body. In particular, let party $j$ value each seat representing ideology $t$ by a standard quadratic utility function:

$$u_j(t) = -(x_j - t)^2$$

Hence, party's $j \in \{L, R\}$ payoff out of the constituted elected body is given by:

$$U_j([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \int_0^1 -(x_j - t)^2 f(t) dt$$

or given the properties of the linear rule and the distribution of ideologies in the elected body according to $f(x)$:

$$U_j([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = s_L([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) \int_{\underline{x}_L}^{\bar{x}_L} -(x_j - t)^2 \frac{1}{\bar{x}_L - \underline{x}_L} dt + s_R([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) \int_{\underline{x}_R}^{\bar{x}_R} -(x_j - t)^2 \frac{1}{\bar{x}_R - \underline{x}_R} dt$$

where given the uniform distribution of voters and the properties of the linear rule $s_L([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \frac{1-n+n(\overline{x}_L+\underline{x}_R)}{2}$ and $s_R([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \frac{1-n+n(2-\overline{x}_L-\underline{x}_R)}{2}$.

Notice that parties' preferences over outcomes can be micro-founded following a logic similar to that of the legislative bargaining model (agenda-setting) of Romer and Rosenthal (1978, 1979) where each legislator is randomly recognized with an equal probability (uniform probability distribution) to make a take it or leave it proposal. In our setup, the Romer-Rosenthal logic suggests that our utility specification above can have a similar interpretation, so that an elected legislator is randomly selected to make a proposal.[2] Thus, parties face a trade-off between ideological dispersion of their elected candidates (preference for a narrower list) and the number of their legislators (i.e., seats) –which is increasing in their vote share, itself increasing with more list dispersion– so as to increase the probability that one of their own legislators is being recognized as the proposer.

**Equilibrium:** Given the presented model, we can now provide the formal result

[2]Interpreting the policy formation stage in our model as a random draw of an ideology of a council member seems compatible with Finnish municipal elections where the role of parties as legislative teams is weak and individual members do affect policy (Hyytinen et al. 2018a; Harjunen et al. 2023; Meriläinen 2022). In this setting, our voters are not less sophisticated than those in standard electoral competition models with expressive voters since despite being non strategic still internalize, at least in part, the policy formation process: they understand that a council member —not a party— will eventually be chosen to implement their ideal policy. So voting for the candidate that better matches their preferences is an uncomplicated way of expressing them. This is exactly the way expressive voting works in winner-take-all electoral competition models (see, e.g., Osborne and Slivinski 1996): candidates locate on the policy space, voters vote according to their preference ranking —ignoring their effect on outcomes, and given the vote profile and the electoral rule, which might be deterministic (see, e.g., Matakos et al. 2016) or probabilistic (Iaryczower and Mattozzi 2013), a candidate is chosen to implement their policy.

characterizing the equilibrium of our model.

**Proposition 1.** *Let* $x^* = \frac{(3+4D)x_L - 3x_L^2 - 1}{x_L + 4D - 1}$. *There exists a unique symmetric equilibrium where:*

1. $[\underline{x}_L^*, \overline{x}_L^*] = [(3x_L - min\{x^*, 0.5\})/2, min\{x^*, 0.5\}]$
2. $[\underline{x}_R^*, \overline{x}_R^*] = [1 - min\{x^*, 0.5\}, 1 - (3x_L - min\{x^*, 0.5\})/2]$
3. *In equilibrium, parties' ideological heterogeneity* $(\overline{x}_j^* - \underline{x}_j^*)$ *is increasing in the electoral rule disproportionality. Formally,* $\frac{\partial(\overline{x}_j^* - \underline{x}_j^*)}{\partial D} \geq 0$, *for both* $j = L, R$.

In the unique symmetric equilibrium, parties' optimal levels of intraparty ideological heterogeneity are given by two equal length intervals on the left and on the right of the median. These are characterized in parts 1 and 2 of the Proposition. Each party $j \in \{L, R\}$ strategically chooses how far from its ideal point its candidates' list should extend depending on disproportionality of the electoral rule $D$ and its ideal point $(x_j)$. Indeed, larger values of $D$ indicate a more disproportional allocation of seats in favor of the larger party.

The comparative static result in part 3 of the Proposition shows that the length of the list is increasing as the rule favors disproportionally the winner of the election (i.e., $\partial(\overline{x}_j^* - \underline{x}_j^*)/\partial D \geq 0$), and is the one presented in our main text as Prediction 1.

The above characterization is the one illustrated in Figure 3. The *min* operator appearing in the formal result simply restricts the equilibrium values $[\underline{x}_j, \overline{x}_j]$ in the admissible strategy space but does not add any essential dynamics to the presented story. As also illustrated in Figure 3, once the most moderate candidates of the lists hit the 0.5 bound, then parties stop including more extreme candidates in their list.

**Proof of Proposition 1:** To provide a unified analysis without repeating the arguments for a number of corner scenarios, throughout this proof we assume that parties can even propose degenerate intervals (we will establish that this never happens in equilibrium) and slightly abuse notation by considering that $\int_a^a -(x-t)^2 \frac{1}{a-a} dt =$

$\lim_{b\to a^+}\int_a^b -(x-t)^2\frac{1}{b-a}dt = -(x-a)^2$. Hence, for every admissible strategy pair, the utility of party $L$ is given by:

$$U_L([\underline{x}_L,\bar{x}_L],[\underline{x}_R,\bar{x}_R]) =$$
$$s_L(\tfrac{\bar{x}_L+\underline{x}_R}{2})\int_{\underline{x}_L}^{\bar{x}_L} -(x_L-t)^2\frac{1}{\bar{x}_L-\underline{x}_L}dt + [1-s_L(\tfrac{\bar{x}_L+\underline{x}_R}{2})]\int_{\underline{x}_R}^{\bar{x}_R} -(x_L-t)^2\frac{1}{\bar{x}_R-\underline{x}_R}dt.$$

Notice that, for every $\bar{x}_L$ and $\underline{x}_R$ such that $s_L(\frac{\bar{x}_L+\underline{x}_R}{2}) > 0$, we have $\frac{\partial^2 U_L([\underline{x}_L,\bar{x}_L],[\underline{x}_R,\bar{x}_R])}{\partial \underline{x}_L^2} = -\frac{2}{3}s_L(\frac{\bar{x}_L+\underline{x}_R}{2}) < 0$ and $\frac{\partial U_L([\underline{x}_L,\bar{x}_L],[\underline{x}_R,\bar{x}_R])}{\partial \underline{x}_L} = \frac{1}{3}s_L(\frac{\bar{x}_L+\underline{x}_R}{2})(3x_L - 2\underline{x}_L - \bar{x}_L)$. Notice that the sign of this derivative is independent of the strategy of $R$ and only depends on the strategy of $L$ and, hence, the unique $\underline{x}_L$ that maximizes the utility of $L$ is a function of $\bar{x}_L$, which we denote by $\underline{x}_L^*(\bar{x}_L)$. Moreover, when $\bar{x}_L$ and $\underline{x}_R$ are such that $s_L(\frac{\bar{x}_L+\underline{x}_R}{2}) = 0$, then, trivially, $\underline{x}_L^*(\bar{x}_L)$ still maximizes the utility of $L$.[3] So, for every fixed strategy of $R$, the problem of $L$ reduces to just selecting the $\bar{x}_L$ that maximizes $U_L([\underline{x}_L^*(\bar{x}_L),\bar{x}_L],[\underline{x}_R,\bar{x}_R])$. By a symmetric argument, we have that the $\bar{x}_R$ that maximizes the utility of $R$ for a fixed triplet $\underline{x}_L$, $\bar{x}_L$, and $\underline{x}_R$ is a function only of $\underline{x}_R$, which we denote by $\bar{x}_R^*(\underline{x}_R)$.

It is easy to see that, for every admissible strategy of $R$ the $\bar{x}_L$ that maximizes $U_L([\underline{x}_L^*(\bar{x}_L),\bar{x}_L],[\underline{x}_R,\bar{x}_R])$ must be at least as large as $x_L$. To see this, consider on the contrary that $U_L([\underline{x}_L^*(\bar{x}_L),\bar{x}_L],[\underline{x}_R,\bar{x}_R])$ is maximized at some $\bar{x}'_L < x_L$. Obviously, $\bar{x}'_L$ must be such that $s_L(\frac{\bar{x}_L+\underline{x}_R}{2}) > 0$. Indeed, $U_L([x_L,\frac{1}{2}],[\underline{x}_R,\bar{x}_R])$, for example, induces $s_L(\frac{\frac{1}{2}+\underline{x}_R}{2}) > 0$ and it is strictly larger than $\int_{\underline{x}_R}^{\bar{x}_R} -(x_L-t)^2\frac{1}{\bar{x}_R-\underline{x}_R}dt$. Hence, $\bar{x}'_L$ cannot be such that $L$ does not elect representatives in the elected body. If $L$ deviates to proposing only candidates with an ideal policy almost identical to $x_L$ (that is, to $[x_L-\epsilon, x_L+\epsilon]$), for any arbitrarily small $\epsilon \in (0, x_L-\bar{x}'_L)$, then $L$ is strictly better off since it elects more members (because $s_L(\frac{\bar{x}'_L+\underline{x}_R}{2}) < s_L(\frac{x_L+\underline{x}_R}{2})$) and all the elected members are better according to its policy preferences. This means that we can focus attention on the restricted form of the game in which players just select their most moderate end of their list from policies at most as extreme as their ideal policies.

[3] The fact that there might be many admissible values of $\underline{x}_L$ that minimize $U_L([\underline{x}_L,\bar{x}_L],[\underline{x}_R,\bar{x}_R])$ when $s_L(\frac{\bar{x}_L+\underline{x}_R}{2}) = 0$ does not pose any threat to our equilibrium uniqueness arguments, since in a symmetric equilibrium $s_L(\frac{\bar{x}_L+\underline{x}_R}{2}) = \frac{1}{2}$.

When $\bar{x}_L \geq x_L$, we have that $\underline{x}_L^*(\bar{x}_L) = \frac{3}{2}(x_L - \frac{\bar{x}_L}{3}) \in (0, x_L)$ and, similarly, when $\underline{x}_R \leq x_R$, we have that $\bar{x}_R^*(\underline{x}_R) = \frac{3}{2}(1 - x_L - \frac{\underline{x}_R}{3}) \in (1 - x_L, 1)$, which implies that

$$\int_{\underline{x}_L^*(\bar{x}_L)}^{\bar{x}_L} -(x_L - t)^2 \frac{1}{\bar{x}_L - \underline{x}_L^*(\bar{x}_L)} dt = -\frac{1}{4}(x_L - \bar{x}_L)^2$$

and that

$$\int_{\underline{x}_R}^{\bar{x}_R^*(\underline{x}_R)} -(x_L - t)^2 \frac{1}{\bar{x}_R^*(\underline{x}_R) - \underline{x}_R} dt = \frac{1}{4}[-3 - 13x_L^2 - \underline{x}_R^2 + 2x_L(6 + \underline{x}_R)].$$

Hence,

$$U_L([\underline{x}_L^*(\bar{x}_L), \bar{x}_L], [\underline{x}_R, \bar{x}_R^*(\underline{x}_R)]) =$$
$$s_L(\tfrac{\bar{x}_L + \underline{x}_R}{2})(-\tfrac{1}{4}(x_L - \bar{x}_L)^2) + [1 - s_L(\tfrac{\bar{x}_L + \underline{x}_R}{2})]\tfrac{1}{4}[-3 - 13x_L^2 - \underline{x}_R^2 + 2x_L(6 + \underline{x}_R)].$$

By linearity of $s_L(\frac{\bar{x}_L + \underline{x}_R}{2})$ when it takes values in $(0, 1)$, it follows that $U_L([\underline{x}_L^*(\bar{x}_L), \bar{x}_L], [\underline{x}_R, \bar{x}_R^*(\underline{x}_R)])$ is quasiconcave in $\bar{x}_L$ for any $\underline{x}_R \in [\frac{1}{2}, x_R]$ and, hence by Debreu (1952), this game has a pure strategy equilibrium. Moreover, if

$$\frac{\partial U_L([\underline{x}_L^*(\bar{x}_L), \bar{x}_L], [\underline{x}_R, \bar{x}_R^*(\underline{x}_R)])}{\partial \bar{x}_L}\Big|_{\bar{x}_L = \bar{x}'_L, \underline{x}_R = \underline{x}'_R} = 0$$

and

$$\frac{\partial U_R([\underline{x}_L^*(\bar{x}_L), \bar{x}_L], [\underline{x}_R, \bar{x}_R^*(\underline{x}_R)])}{\partial \underline{x}_R}\Big|_{\bar{x}_L = \bar{x}'_L, \underline{x}_R = \underline{x}'_R} = 0$$

we have an interior equilibrium at $(\bar{x}'_L, \underline{x}'_R)$. We notice that

$$\frac{\partial U_L([\underline{x}_L^*(\bar{x}_L), \bar{x}_L], [\underline{x}_R, \bar{x}_R^*(\underline{x}_R)])}{\partial \bar{x}_L}\Big|_{\bar{x}_L = \bar{x}_L^*, \underline{x}_R = 1 - \bar{x}_L^*} = 0 \Longrightarrow \bar{x}_L^* = \frac{(3 + 4D)x_L - 3x_L^2 - 1}{x_L + 4D - 1}$$

with $\underline{x}_L^*(\bar{x}_L^*) = \frac{3}{2}(x_L - \frac{\bar{x}_L^*}{3})$. Since, $\bar{x}_L^* = \frac{(3+4D)x_L - 3x_L^2 - 1}{x_L + 4D - 1} < \frac{1}{2}$ if and only if $D \in [0, \frac{3x_L - 1}{4})$, we conclude that there is a unique symmetric equilibrium, $([\underline{x}_L^*, \bar{x}_L^*], [\underline{x}_R^*, \bar{x}_R^*]) = ([\underline{x}_L^*, \bar{x}_L^*], [1 - \bar{x}_L^*, 1 - \underline{x}_L^*])$, such that $\bar{x}_L^* = \frac{(3+4D)x_L - 3x_L^2 - 1}{x_L + 4D - 1}$ if $D \in [0, \frac{3x_L - 1}{4})$ and such that $\bar{x}_L^* = \frac{1}{2}$ if $D \geq \frac{3x_L - 1}{4}$. This equilibrium is presented in parts 1 and 2 of Proposition 1 using the min operator. Part 3 of the Proposition stating that $\frac{\partial(\bar{x}_j^* - \underline{x}_j^*)}{\partial D} \geq 0$ holds since $\frac{\partial \frac{(3+4D)x_L - 3x_L^2 - 1}{x_L + 4D - 1}}{\partial D} \geq 0$.□

### A1.1 Discussion of model and robustness

The mechanism we put forward and the resulting prediction (that is, as the electoral rule becomes more disproportional parties become less cohesive) essentially depends on the following premises: First, the personal vote dimension is present, but not necessarily the unique relevant determinant of voters' behavior. Second, ceteris paribus, parties prefer to nominate candidates who share the parties' ideology. This means that adaptations of the model that satisfy these two assumptions should result in comparative statics that are qualitatively similar to those of the model analyzed in the paper while any variation of the model violating any of these assumptions is not guaranteed to produce similar results. There is also a third premise that is key, which we nonetheless do not consider to be a limiting assumption: That parties play an important role in candidate selection. Effectively, opening the black box of intraparty dynamics in determining selection (see e.g., Dal Bò and Finan 2018) is part of our contribution. A detailed discussion of relevant modifications follows.

#### Voting behavior

So far we assumed that voters vote for the party that included in its list the candidate closest to their ideal point, that is, purely on the basis of a personal vote. Here, we provide two natural extensions. First, we assume that voters may *also* care about the lists' mean ideology. That is, they may care not only for the closest candidate to them but also about the composition of the entire list. Second, we assume that while some voters may care only about the closest candidate to them, others may vote for the list that offers them higher expected utility. Our main comparative static result is robust to these extensions. Moreover, these results emphasize the importance of having (some) personal vote dimension in our model –and, in fact, in any model that aspires to study intra-party ideological heterogeneity– since in the extreme cases where voters exclusively prioritize either the parties' mean ideology or the composition of the parliament our findings qualitatively resemble those of Matakos et al. (2016), leading parties to propose a singleton.

Given proposed platforms $([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R])$ recall that party's $j$ payoff is

$$U_j([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = s_L([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) \int_{\underline{x}_L}^{\bar{x}_L} -(x_j - t)^2 \frac{1}{\bar{x}_L - \underline{x}_L} dt + s_R([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) \int_{\underline{x}_R}^{\bar{x}_R} -(x_j - t)^2 \frac{1}{\bar{x}_R - \underline{x}_R} dt$$

While the seat shares $s_L([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R])$ and $s_R([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R])$ need to be calculated separately for each of the two variations of our original model, the corresponding payoffs for party $j$ from each of the two lists are equal to

$$\int_{\underline{x}_L}^{\bar{x}_L} -(x_j - t)^2 \frac{1}{\bar{x}_L - \underline{x}_L} dt = \frac{-3x_j^2 - \underline{x}_L^2 - \underline{x}_L \overline{x}_L - \overline{x}_L^2 + 3x_j(\underline{x}_L + \overline{x}_L)}{3}$$

and

$$\int_{\underline{x}_R}^{\bar{x}_R} -(x_j - t)^2 \frac{1}{\bar{x}_R - \underline{x}_R} dt = \frac{-3x_j^2 - \underline{x}_R^2 - \underline{x}_R \overline{x}_R - \overline{x}_R^2 + 3x_j(\underline{x}_R + \overline{x}_R)}{3}$$

i) In the first variation of the voting behavior we assume that voters assign weight $\alpha \in (0, 1)$ to the location of each list's mean candidate and $(1-\alpha)$ to each list's candidate closest to their ideal policy. The location of the indifferent voter now determining vote shares and hence seat shares is at $\frac{[\alpha \frac{\underline{x}_L + \overline{x}_L}{2} + (1-\alpha)\overline{x}_L] + [\alpha \frac{\underline{x}_R + \overline{x}_R}{2} + (1-\alpha)\underline{x}_R]}{2}$. Given the uniform distribution of voters and the properties of the linear rule we have that parties' seat shares are equal to

$$s_L([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \frac{1-n}{2} + n \frac{[\alpha \frac{\underline{x}_L + \overline{x}_L}{2} + (1-\alpha)\overline{x}_L] + [\alpha \frac{\underline{x}_R + \overline{x}_R}{2} + (1-\alpha)\underline{x}_R]}{2}$$

and

$$s_R([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \frac{1-n}{2} + n\left(1 - \frac{[\alpha\frac{\underline{x}_L+\overline{x}_L}{2} + (1-\alpha)\overline{x}_L] + [\alpha\frac{\underline{x}_R+\overline{x}_R}{2} + (1-\alpha)\underline{x}_R]}{2}\right).$$

Substituting the above in party's $j$ payoff, taking first order conditions for party $j$ with respect to $\underline{x}_j$ and $\overline{x}_j$ and solving the system of equations assuming a symmetric equilibrium we obtain a unique interior solution $([\underline{x}^*_L, \overline{x}^*_L], [\underline{x}^*_R, \overline{x}^*_R]) = ([\underline{x}^*_L, \overline{x}^*_L], [1-\overline{x}^*_L, 1-\underline{x}^*_L])$ where

$$\underline{x}^*_L = \frac{2 - 3\alpha(1-2x_L)^2 + 4x_L(-3+4D+3x_L)}{4(4D+x_L-1)}$$

and

$$\overline{x}^*_L = \frac{-4 + 3\alpha(1-2x_L)^2 + 4x_L(3+4D-3x_L)}{4(4D+x_L-1)}.$$

Hence, each party's equilibrium level of intraparty cohesion in the interior equilibrium (i.e., for $D$ sufficiently small) is equal to

$$\overline{x}^*_j - \underline{x}^*_j = \frac{3(\alpha-1)(1-2x_L)^2}{2(4D+x_L-1)}$$

implying that parties become less cohesive as the electoral disproportionality increases since

$$\frac{\partial(\overline{x}^*_j - \underline{x}^*_j)}{\partial D} = \frac{6(1-\alpha)(1-2x_L)^2}{(4D+x_L-1)^2} > 0.$$

ii) In the second variation of the voting behavior we assume that a fraction $\alpha \in (0,1)$ of voters with ideal policies distributed uniformly over the policy space compares the expected utility from each party using quadratic losses, while a fraction $(1-\alpha)$ of voters votes for the list including the candidate closest to their ideal policy as in our benchmark model.

In order to calculate parties' seat shares we need to calculate the indifferent voter in each of the two fractions of the population. The indifferent voter among voters in fraction $\alpha$ of the population is located at $(\overline{x}_L + \underline{x}_R)/2$. The indifferent voter among voters in fraction $\alpha$ of the population is voter $i$ with ideal point $x_i$ for whom

$$\int_{\underline{x}_L}^{\bar{x}_L} -(x_i - t)^2 \frac{1}{\bar{x}_L - \underline{x}_L} dt = \int_{\underline{x}_R}^{\bar{x}_R} -(x_i - t)^2 \frac{1}{\bar{x}_R - \underline{x}_R} dt$$

which requires that

$$\frac{-3x_i^2 - \underline{x}_L^2 - \underline{x}_L \overline{x}_L - \overline{x}_L^2 + 3x_i(\underline{x}_L + \overline{x}_L)}{3} = \frac{-3x_i^2 - \underline{x}_R^2 - \underline{x}_R \overline{x}_R - \overline{x}_R^2 + 3x_i(\underline{x}_R + \overline{x}_R)}{3}$$

and hence the location of the indifferent voter is

$$x_i = \frac{\underline{x}_L^2 + \underline{x}_L \overline{x}_L + \overline{x}_L^2 - \underline{x}_R^2 - \underline{x}_R \overline{x}_R - \overline{x}_R^2}{3(\underline{x}_L + \overline{x}_L - \underline{x}_R - \overline{x}_R)}$$

Given the uniform distribution of voters the vote share of party $L$ is hence:

$$v_L([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \alpha \frac{\underline{x}_L^2 + \underline{x}_L \overline{x}_L + \overline{x}_L^2 - \underline{x}_R^2 - \underline{x}_R \overline{x}_R - \overline{x}_R^2}{3(\underline{x}_L + \overline{x}_L - \underline{x}_R - \overline{x}_R)} + (1-\alpha)\frac{\overline{x}_L + \underline{x}_R}{2}$$

Given the properties of the linear rule we have that parties' seat shares are equal to:

$$s_L([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \frac{1-n}{2} + n\left(\alpha \frac{\underline{x}_L^2 + \underline{x}_L \overline{x}_L + \overline{x}_L^2 - \underline{x}_R^2 - \underline{x}_R \overline{x}_R - \overline{x}_R^2}{3(\underline{x}_L + \overline{x}_L - \underline{x}_R - \overline{x}_R)} + (1-\alpha)\frac{\overline{x}_L + \underline{x}_R}{2}\right)$$

and

$$s_R([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \frac{1-n}{2} + n\left[1 - \left(\alpha \frac{\underline{x}_L^2 + \underline{x}_L \overline{x}_L + \overline{x}_L^2 - \underline{x}_R^2 - \underline{x}_R \overline{x}_R - \overline{x}_R^2}{3(\underline{x}_L + \overline{x}_L - \underline{x}_R - \overline{x}_R)} + (1-\alpha)\frac{\overline{x}_L + \underline{x}_R}{2}\right)\right].$$

Substituting the above in party's $j$ payoff, taking first order conditions for party $j$ with respect to $\underline{x}_j$ and $\overline{x}_j$ and solving the system of equations assuming a symmetric equilibrium we obtain a unique interior solution $([\underline{x}_L^*, \bar{x}_L^*], [\underline{x}_R^*, \bar{x}_R^*]) = ([\underline{x}_L^*, \bar{x}_L^*], [1 - \bar{x}_L^*, 1 - \underline{x}_L^*])$ where

$$\underline{x}_L^* = \frac{1}{2} - \frac{(8D-1)(2i-1)(2\alpha i - \alpha - 4D - 3i + 2)}{2(4D + i - 1)(2\alpha i - \alpha + 8D - 1)}$$

and

$$\overline{x}_L^* = \frac{1}{2} + \frac{(8D-1)(2i-1)(4\alpha i - 2\alpha + 4D - 3i + 1)}{2(4D+i-1)(2\alpha i - \alpha + 8D - 1)}.$$

Hence, each party's equilibrium level of intraparty cohesion in the interior equilibrium (i.e., for $D$ sufficiently small) is equal to

$$\overline{x}_j^* - \underline{x}_j^* = \frac{3(\alpha-1)(8D-1)(1-2i)^2}{2(4D+i-1)(2\alpha i - \alpha) + 8D - 1)}$$

where the derivative of the party list when the electoral disproportionality varies is

$$\frac{\partial(\overline{x}_j^* - \underline{x}_j^*)}{\partial D} = \frac{6(1-\alpha)(2i-1)^2\,(4\alpha i - 4\alpha i^2 - \alpha + 64D^2 - 16D + 1)}{(4D+i-1)^2(2\alpha i - \alpha + 8D - 1)^2}$$

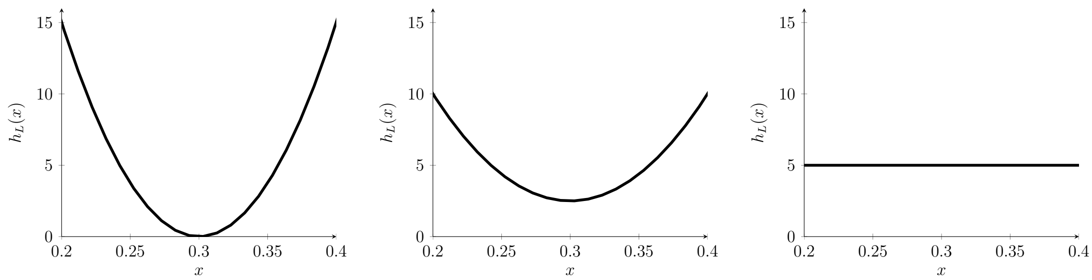
implying that in the presence of enough voters voting for the closest candidate parties become less cohesive as the electoral disproportionality increases (i.e., $\frac{\partial(\overline{x}_j^* - \underline{x}_j^*)}{\partial D} > 0$ for $\alpha < \frac{64D^2-16D+1}{4i^2-4i+1}$). Notice that when disproportionality is not very large (i.e., $D < 0.083$) then $\frac{\partial(\overline{x}_j^* - \underline{x}_j^*)}{\partial D} > 0$ is true for any $0 < \alpha < 1$. Indeed, in our empirical exercise the smallest council is of size $k = 13$ and hence the most disproportional linear rule that approximates the D'Hondt rule of this size is characterized by $D = 1/(8k) = 0.0096 < 0.083$. Moreover, this derivative is positive even when (i.e., $D \geq 0.083$), as long as the share of voters who vote for the closest candidate is sufficiently large (i.e., $\alpha < \frac{64D^2-16D+1}{4i^2-4i+1}$).

### Non uniform distribution

In our benchmark model we assumed that the distribution of ideologies in the elected body has support on the lists chosen by the parties (i.e., $[\underline{x}_L, \overline{x}_L] \cup [\underline{x}_R, \overline{x}_R]$), and ideologies are uniformly distributed within each party. We now relax the assumption of the uniform distribution. Formally, within each party $j$ proposing interval $[\underline{x}_j, \overline{x}_j]$, we assume that the within party distribution of ideologies is given by the following density

$$h_j(x) = \begin{cases} \dfrac{\left[12(1-a)\left(\frac{\overline{x}_j - x}{\overline{x}_j - \underline{x}_j}\right)^2 - 12(1-a)\left(\frac{\overline{x}_j - x}{\overline{x}_j - \underline{x}_j}\right) + (3-2a)\right]}{\overline{x}_j - \underline{x}_j} & \text{if } \underline{x}_j \leq x \leq \overline{x}_j \\ 0 & \text{otherwise} \end{cases}$$

where parameter $a \in [0,1]$ determines the shape of the density function with $h_j((\overline{x}_j + \underline{x}_j)/2) = \frac{a}{\overline{x}_j - \underline{x}_j}$. For $a = 1$ we have a uniform distribution as in our benchmark model. For $a = 0$ we have a quadratic density with the midpoint of the list having a zero weight. Values in between vary the shape of the density moving weight from the extremes to moderate policies. Figure A1 illustrates an example using this density.




(a) $a = 0$ (b) $a = 0.5$ (c) $a = 1$

Figure A1: The distribution of the represented ideologies in the elected body for party $L$ when proposing $[\underline{x}_L, \overline{x}_L] = [0.2, 0.4]$, for different values of $a$.

Given proposed platforms $([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R])$ that result to a seat distribution $(s_L, s_R)$ we now have that the density of ideologies in the elected body are given by the following function:

$$f(x) = \begin{cases} s_L h_L(x) & \text{if } \underline{x}_L \leq x \leq \overline{x}_L \\ s_R h_R(x) & \text{if } \underline{x}_R \leq x \leq \overline{x}_R \\ 0 & \text{otherwise} \end{cases}$$

Recall that party's $j$ payoff is equal to:

$$U_j([\underline{x}_L, \overline{x}_L], [\underline{x}_R, \overline{x}_R]) = \int_0^1 -(x_j - t)^2 f(t) dt$$

where the density $f(t)$ is defined above. Substituting the density in the utility function and following similar steps as in the proof of our benchmark result we can first show that for any $\overline{x}_L$ there is a unique $\underline{x}_L$ that maximizes $L$'s payoff (with similar arguments holding for party $R$). Focusing on party $L$ we have that $\underline{x}_L^*(\bar{x}_L) = \frac{-15x_L+(3+2a)\bar{x}_L}{2(a-6)}$ and the symmetric best response for party $R$. So, same as in our benchmark model, for every fixed strategy of $R$, the problem of $L$ reduces to just selecting the $\bar{x}_L$ that maximizes $U_L([\underline{x}_L^*(\bar{x}_L), \bar{x}_L], [\underline{x}_R, \bar{x}_R])$ and similarly for party $R$.

Taking first order condition for party $L$ with respect to $\overline{x}_L$ and assuming a symmetric equilibrium we obtain a unique interior solution $([\underline{x}_L^*, \bar{x}_L^*], [\underline{x}_R^*, \bar{x}_R^*]) = ([\underline{x}_L^*, \bar{x}_L^*], [1-\bar{x}_L^*, 1-\underline{x}_L^*])$ where

$$\underline{x}_L^* = \frac{a(32Dx_L - 2) + (30 - 72D)x_L - 30x_L^2 - 3}{2(4a-9)(4D + x_L - 1)}$$

and

$$\overline{x}_L^* = \frac{a(16Dx_L - 1) - (36D + 15)x_L + 15x_L^2 + 6}{(4a-9)(4D + x_L - 1)}$$

For low enough levels of disproportionality giving rise to the above equilibrium the intra-party cohesion is equal to:

$$\overline{x}_j^* - \underline{x}_j^* = \frac{15(1-2x_L)^2}{2(4a-9)(4D + x_L - 1)}$$

where it holds that:

$$\frac{\partial(\overline{x}_j^* - \underline{x}_j^*)}{\partial D} = \frac{30(1-2x_L)^2}{(9-4a)(4D + x_L - 1)^2} > 0.$$

and hence our main result holds also in this setting. That is, parties lists are less cohesive as the disproportionality of the electoral rule increases.

### Electoral rules

We presented our formal results using a linear seat allocation method. This is the simplest electoral rule that captures the described incentives as the electoral rule disproportionality varies. So one may reasonably wonder how robust our results are to other electoral rules. Below we first argue that the result extends to other non-linear continuous allocation rules satisfying mild properties. Hence, our results are not an artifact of the linearity assumption. Furthermore, we also solve the model using the D'Hondt rule present in our empirical setting. Hence, our results are not an artifact of the continuity assumption, after all, all electoral rules in reality are by definition discontinuous. Both alterations in the electoral rule show that our main results are robust to non linear rules.

***Theil's rule***: (Theil 1969) suggested that parties' seat share ratio depends on the vote share ratio in a non linear manner. This rule is extensively used in the literature both due to its tractability and the fact that it encompasses the "cube law", a typical way of modeling first majoritarian elections (see Herrera et al. 2014, 2016, 2019; Matakos et al. 2016 for recent applications and Taagepera (1986) on the cube law).

Formally, Theil's rule is allocating seat shares as follows:

$$\frac{s_L}{s_R} = \left(\frac{v_L}{v_R}\right)^n \implies s_L = \frac{v_L^n}{v_L^n + (1 - v_L)^n}$$

where $n \geq 1$. If $n = 1$, parties' seat shares are equal to their vote shares. If $n > 1$, the electoral rule favors the winner of the election. This advantage for the election winner increases in $n$, same as in the case of the linear rule. Figure A2 illustrates a pure PR system ($n = 1$), a a winner-take-all election where the winner is allocated all the seats in the elected body ($n \to \infty$), and an intermediate case with $n = 3$, the so called cube law considered a "standard" approximation of first-past-the-post systems with several districts (Taagepera 1986).

Using Theil's rule instead of the Linear rule we obtain qualitatively identical results to the ones of Proposition 1. Now the expression for $x^*$ is given by

$$x^* = \frac{x_L + [2 + x_L(6x_L - 7)]n}{1 + (1 - 2x_L)n}$$

and given that an increase in $n$ implies a higher disproportionality, we are again obtain-

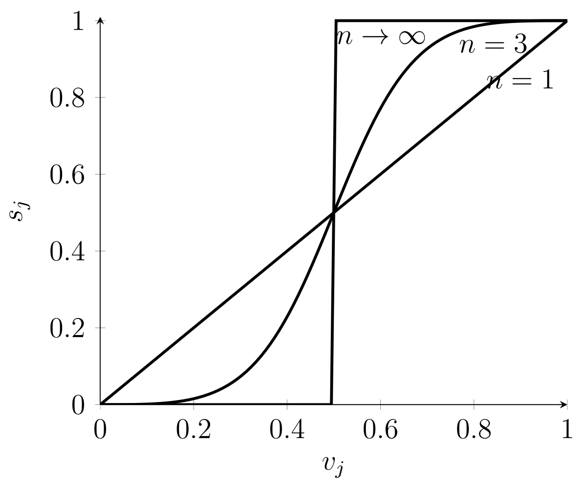


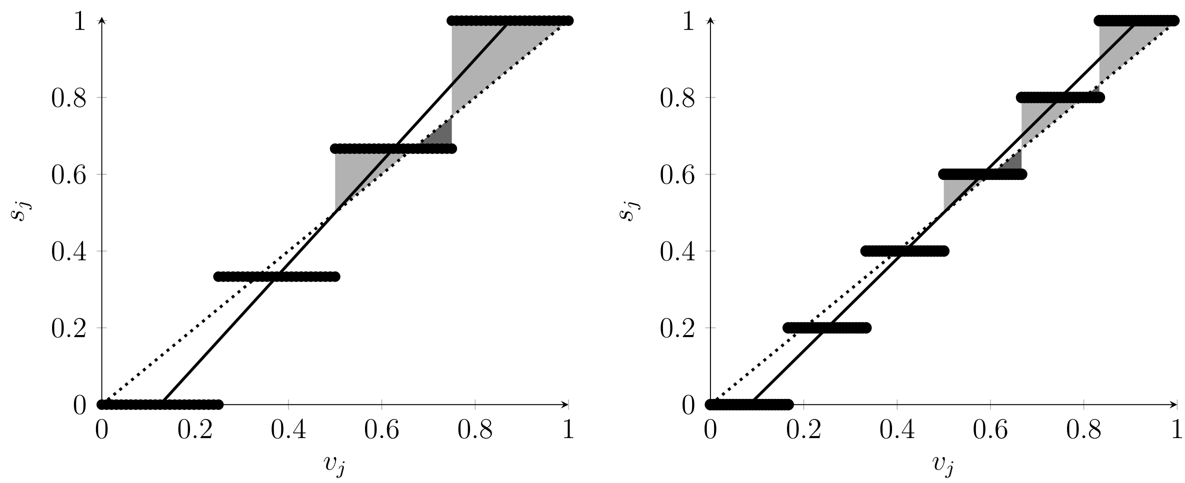

Figure A2: Seat share allocation given parties' vote shares according to ***Theil's rule*** where $s_j(v_j) = \frac{v_j^n}{v_j^n + (1-v_{-j})^n}$

ing our main prediction where parties' lists cohesion is decreasing in the electoral rule disproportionality $D$.

***D'Hondt method:*** In our empirical setting seats are allocated according to the D'Hondt method in all municipalities. The D'Hondt method operates as follows. It assigns the first seat to the most voted party. From then on, given the votes of party $j$ (i.e., $v_j$) and the seats received (so far) by party $j$, (i.e., $s_j$), the method compares the corresponding values of the quotient $v_j/(s_j+1)$ for each party $j$ and assigns the next seat to the party with the highest quotient until all $k$ seats are allocated.

For a two-party example with three seats to be allocated, let the winning party obtain 60 votes and the losing party 40. The method assigns the "first" seat to the winner (since 60>40), the "second" seat to the losing party (comparing the quotients 40>30), and the "third" seat to the winning party (since 30>20). Following this allocation rule, the step function in Figure A3 summarizes the seat allocation in a two-party election when three or five seats are allocated. Figure A3 also makes evident that for a two-party election, each seat is allocated to a party for every $1/(k+1)$ vote share obtained.[4]

[4]Formally, for a two-party election, if $v_j \geq 0.5$, party $j$ obtains $\hat{j} \in \{0, 1, ..., k\}$ out of the $k \in \mathbb{N}^+$ seats where $\hat{j}$ is the unique value such that $\hat{j}/(k+1) \leq v_j < (\hat{j}+1)/(k+1)$. If $v_j < 0.5$, party $j$ obtains $\hat{j}$ out of the $k$ seats where $\hat{j}$ is the unique value such that $\hat{j}/(k+1) < v_j \leq (\hat{j}+1)/(k+1)$. Hence, $s_j = \hat{j}/k$. The distinction between the cases where $v_j \geq 0.5$ or $v_j < 0.5$ is assuming that the tie-breaking




(a) Council size $k = 3$ and "Slope"=1.333 (b) Council size $k = 5$ and "Slope"=1.2

Figure A3: D'Hondt allocation method (steps) in a council of size $k$ and its continuous approximation using the *Linear rule* (solid line) with slope $n = (k+1)/k$ (for the strictly increasing part). Dotted line is the 45 degree line (i.e., pure PR for $k \to \infty$). Light (dark) gray areas denote vote shares where the winner of the election is favored (harmed) by the allocation method.

Figure A3 captures for a two-party scenario: a) how the D'Hondt rule is in expectation favoring the winner of the election,[5] b) how the electoral rule disproportionality (i.e., the expected advantage of the winner) is decreasing in the council size (the key element for our identification),[6] and c) how the *Linear rule* presented in our conceptual framework links with our empirical analysis as a linear approximation of the discontinuous D'Hondt method (Flis et al. 2020).[7] In reality, Finland has a multi-party system and,

rule is favoring the winner of the election. One can permit different tie breakers as long as the symmetry assumption is guaranteed.

[5]The seat allocation following the step function deviates from the 45 degree line representing a seat share equal to the vote share. While there are realizations of vote shares that the winner is harmed by the seat allocation (dark grey areas), for most realizations the winner is favored by the seat allocation (light gray areas). Both panels illustrate how the total light gray area is larger than the total dark great area. Hence, the election winner is expected to be favored by the electoral rule. This implies that our measure of disproportionality $D$ for the D'Hondt method, measuring deviations of the seat share from the 45 degrees line for $v_j > 50\%$, is strictly positive.

[6]Our measure of disproportionality $D$ is decreasing in the council size $k$. That is made evident on the graph for a council of size 3 and 7. Analytically, the value of the electoral rule disproportionality in a two-party election under the D'Hondt system is $D = 1/(8k)$, which is strictly decreasing in the council size $k$. This value is obtained by calculating $D = \sum_{j=0}^{\frac{k+1}{2}-1}[\frac{1}{2}(\frac{\frac{k+1}{2}+j}{k} - \frac{\frac{k+1}{2}+j}{k+1})^2 - \frac{1}{2}(\frac{\frac{k+1}{2}+j+1}{k+1} - \frac{\frac{k+1}{2}+j}{k})^2]$ which corresponds to the summation of the light gray areas minus the dark gray areas presented in Figure A3.

[7]The solid line that passes through the midpoint of each step is the *Linear rule* assuming its slope is $n = (k+1)/k$, now depending on the council size. As the council size $k$ increases, the line becomes flatter and, therefore, the rule less disproportional, with $k \to \infty$ leading to a pure PR rule. On more

as well known, the number of parties is an essential feature of different electoral systems (Duverger 1954).[8] From a theoretical perspective, the arguments made regarding the D'Hondt rule, which are expected to favor parties obtaining high vote shares, are also valid in a multiparty context (Gallagher 1991; Herron et al. 2018; Schuster et al. 2003). As Schuster et al. (2003) formally show, *"as the number of parties grows, the deficits of the smaller parties accumulate and generate a growing surplus of seats for the larger parties"*. Perhaps most importantly, in Section 3.4 we precisely provide causal evidence showing that in Finish multiparty municipal elections, the rule's disproportionality is decreasing in the council size.

Given the discontinuity introduced by the D'Hond rule, same as for any rule allocating a discrete number of seats, an equilibrium characterization proves less tractable than in our original model. We therefore illustrate the main comparative statics result by solving our model assuming a discrete policy space. Figure A4 presents this result assuming that the policy space consists of 501 equidistant discrete locations in $[0, 1]$ and that parties ideal policies are at 0.4 and 0.6.

The equilibrium is now obtained in mixed strategies, with parties assigning different weights to different policy intervals. Figure A4 plots the expected list width for different number of seats to be allocated (seats in line with councils in our empirical setting). The illustrated result is in line with our main argument: As the council size increases (i.e., the rule becomes less disproportional —see arguments in Section 3.1), parties propose in expectation narrower lists.[9]

details how a linear function serves as a good continuous approximation of the D'Hondt rule suitable for tractable (empirical) analysis see Flis et al. (2020).

[8] Currently, there are eight parties in the Finnish parliament and these same parties also dominate municipal politics, but some local single-issue groups exist as well. For example, in the 2008 municipal elections the three largest parties (the Social Democrats, the Centre Party and the National Coalition) received around 65 percent of the votes with roughly similar overall shares but with large variation in shares between municipalities.

[9] Note that the jumps observed in the graph are due to discreteness of the policy space. Finer policy spaces can generate an even smoother decreasing curve.

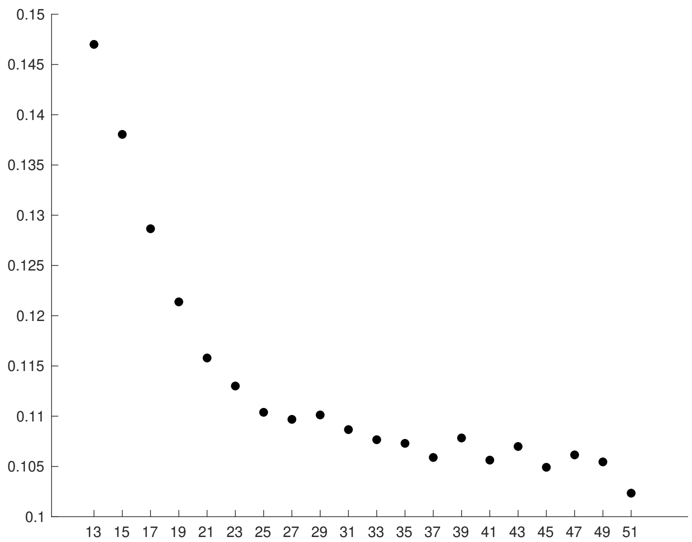



Figure A4: Parties' expected list width as a function of council size solving our model with a discrete policy space and the D'Hondt rule. In the graph, we assume 501 equidistant discrete locations in $[0, 1]$ and that parties ideal policies are at 0.4 and 0.6.

### Number of parties

As almost any model of electoral competition, an equilibrium characterization may be challenging if we were to permit more than two parties. However, we can illustrate how the main trade-off that parties face when choosing their lists should also be present in multiparty settings. While a full equilibrium analysis of a multiparty scenario is intractable without over-simplifying other assumptions substantially, the trade-offs that drive the comparative results clearly do not hinge on a particular number of parties involved. After all, if a more inclusive party list increases the electoral performance but is unappealing policy-wise, then whenever the electoral rule rewards an increase in the vote share more, the party should expand its list, independently of how many competitors it faces. Thus, our formal results need to be thought as a comparative static of the electoral rule disproportionality on intraparty cohesion for a *fixed number of parties*. This situation is in line with our empirical results, as detailed in Section 3.6 where there is no significant jump in the number of competing parties when crossing the threshold.

To see why our main arguments do not hinge on the presence of exactly two parties, consider, for instance, a set of three parties, $M = \{1, 2, 3\}$ and define by $s_i(v)$ the seat share of party $i \in M$ when the distribution of vote shares is given by $v = (v_1, v_2, v_3)$. Then, the utility of party $i$ when each party $j \in M$ proposes a list $[a_j, b_j]$ is given by:

$$\sum_{j \in M} s_j(v) \int_{a_j}^{b_j} -\frac{1}{b_j - a_j}(x_i - t)^2 dt$$

where $x_i$ is simply the ideal policy of party $i$. If $x_2 = 1/2$ and we are in a symmetric situation (i.e., $v_1 = v_2 = v_3$), with $[a_2, b_2] = [1/2 - d, 1/2 + d]$, $[a_1, b_1] = [1 - b_3, 1 - a_3]$ and non-overlapping lists, then, the marginal gain of the moderate party from expanding its list (i.e., from increasing $d$) is equal to:

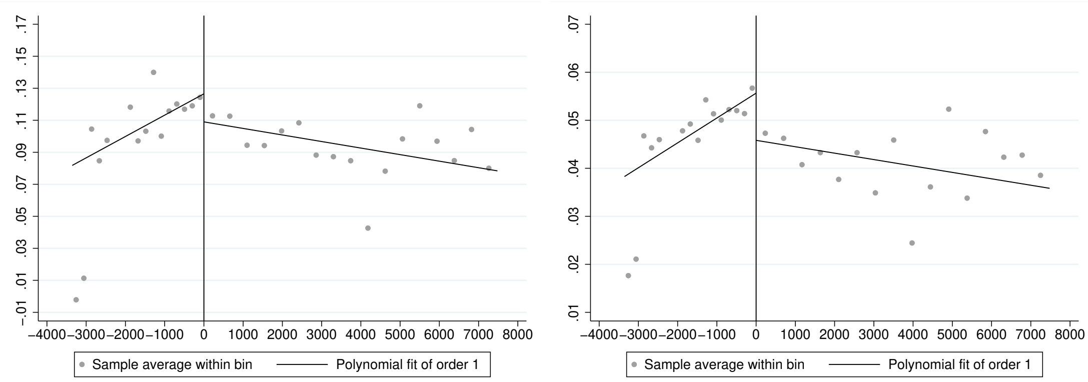
$$\sum_{j \in M} \frac{\partial s_j(v)}{\partial d} \int_{a_j}^{b_j} -\frac{1}{b_j - a_j}(\tfrac{1}{2} - t)^2 dt - \tfrac{2}{3} d s_2(v).$$

Considering that the electoral rule is anonymous, we must have $s_j(v) = \frac{1}{3}$ for every $j \in M$, $\sum_{j \in M} \frac{\partial s_j(v)}{\partial d} = 0$ and $2\frac{\partial s_1(v)}{\partial d} = 2\frac{\partial s_3(v)}{\partial d} = -\frac{\partial s_2(v)}{\partial d} < 0$. If the rule changes from $s$ to some other anonymous rule, $\hat{s}$, with $\frac{\partial \hat{s}_2(v)}{\partial d} > \frac{\partial s_2(v)}{\partial d}$ (i.e., if the rule now rewards more an increase in vote shares compared to the old one), the marginal gain of the moderate party from expanding its list becomes unambiguously larger and, therefore, a more inclusive list becomes more appealing than before. Indeed, similar arguments hold true for the extremist parties as well and, hence, the intuition of the detailed equilibrium analysis provided qualifies to more general setups.

Of course, a fixed number of parties may seem at first a strong assumption when thinking of different electoral rules. Dating back to Duverger (1954), PR systems leading to a multiparty system and first past the post (FPTP) systems leading to a two-party system are considered a “law”, and there is indeed supportive evidence (e.g., Bol et al. 2019 and references therein). However, pure PR and FPTP systems are just the two extreme cases of our model. In between these two extremes, we permit variations in the electoral rule disproportionality. These marginal changes in disproportionality may not be large enough to lead to a change in the number of parties, a likely situation in our empirical setting (see Section 3.6)

## A2 Robustness and validity checks

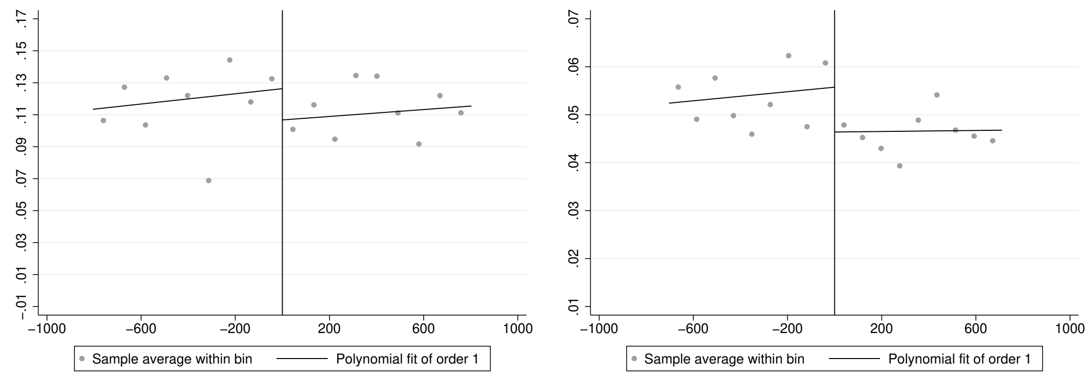
First, we analyze whether the visual inspection of the jump at the cutoff presented in Figures 5 and 7 and is robust to using a **linear fit** instead of the fourth order polynomial. We conduct the analysis for all our realized distortion and cohesion measures, and both for the whole data and for limiting the sample within the MSE-optimal bandwidths. These are shown in Figures A5, A6, A7 and A8.




(a) Slope index.

(b) Modified Gallagher index.

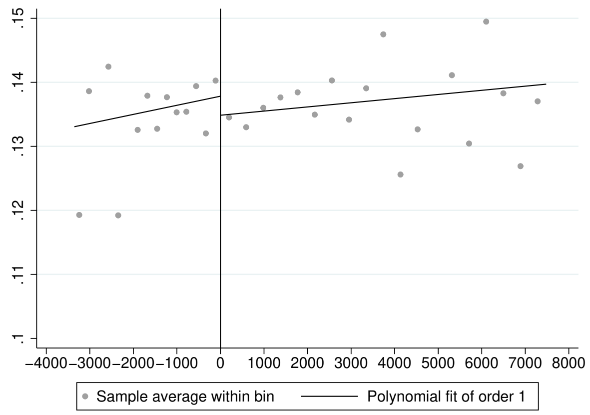
Figure A5: Pooled RDD. We use rdplot package in STATA. We report bins that mimic variance by using evenly-spaced method with spacings estimators. We use $1^{st}$ order polynomial for the fit.



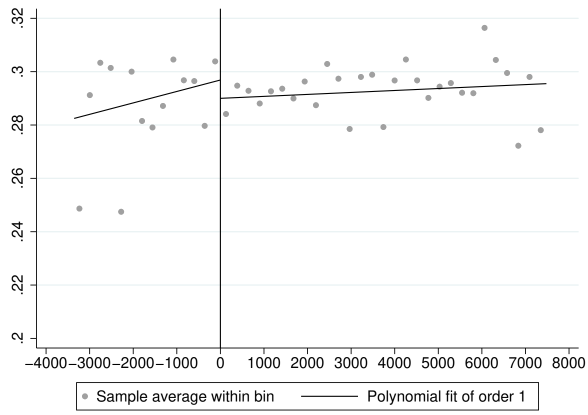

(a) Slope index.

(b) Modified Gallagher index.

Figure A6: Pooled RDD. We use rdplot package in STATA. We report bins that mimic variance by using evenly-spaced method with spacings estimators. We use $1^{st}$ order polynomial for the fit. Sample is limited within the MSE-optimal bandwidth.

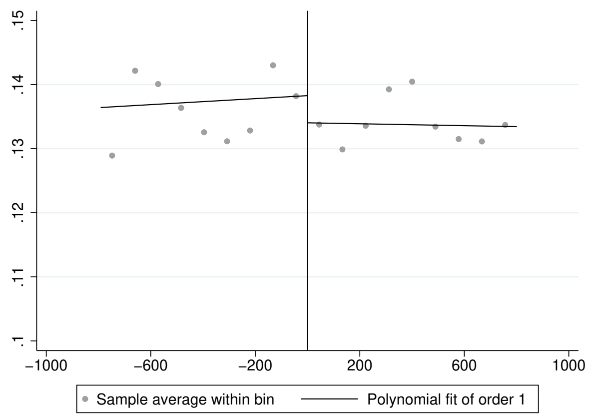



(a) All questions index.

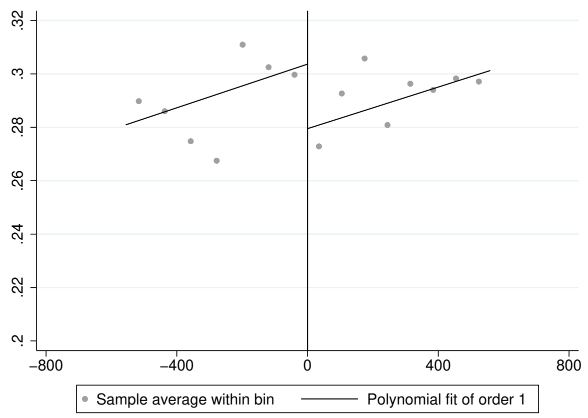



(b) Redistribution index.

Figure A7: Pooled RDD. We use rdplot package in STATA. We report bins that mimic variance by using evenly-spaced method with spacings estimators. We use $1^{st}$ order polynomial for the fit.




(a) All questions index.




(b) Redistribution index.

Figure A8: Pooled RDD. We use rdplot package in STATA. We report bins that mimic variance by using evenly-spaced method with spacings estimators. We use $1^{st}$ order polynomial for the fit. Sample is limited within the MSE-optimal bandwidth.

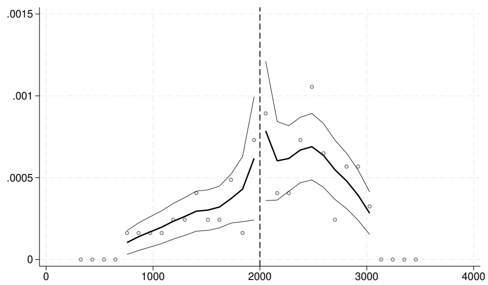
Second, in Table A1, we show that the main cohesion results are robust to using a more flexible specification where we include cutoff specific fixed effects, and allow different linear trends in the running variable around each cutoff and optimize the bandwidths for each cutoffs separately before pooling the data. Notice that due to using a richer model we cannot use the rdrobust-package, and thus, only report conventional inference.

Table A1: More flexible policy position RDD

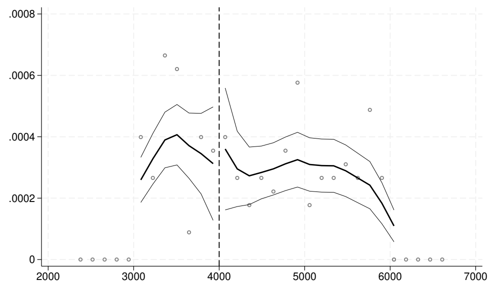
| | **All questions** | **All questions** | **Redistribution** | **Redistribution** |
|---|---|---|---|---|
| Conventional local linear RD coefficient | -0.0092 | -0.0092 | -0.0180 | -0.0176 |
| 95% confidence interval | [-0.0174 ; -0.0001] | [-0.0174 ; -0.0001] | [-0.0361 ; 0.0001] | [-0.0390 ; 0.0038] |
| N within bandwidth | 2529 | 2529 | 2745 | 2745 |
| Clustered bandwidths and s.e.'s | Yes | Yes | Yes | Yes |
| Cutoff fixed effects | Yes | Yes | Yes | Yes |
| Cutoff specific trends | No | Yes | No | Yes |

Notes: Results are generated using using MSE-optimal bandwidth optimized for each cutoff separately before pooling the cutoffs, and a triangular kernel. We use local linear regression to construct the point estimator. Sample is limited to municipalities with population below 22500. Unit of observation is a candidate. Standard errors are clustered at the municipality level.

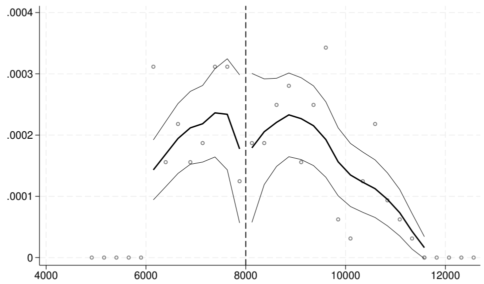
Third, in Figure A9, we present the **McCrary (2008) test for manipulation** separately for each threshold as well as for the pooled data. We use the DcDensity package in STATA. There is no indication of municipalities sorting across the cutoffs.



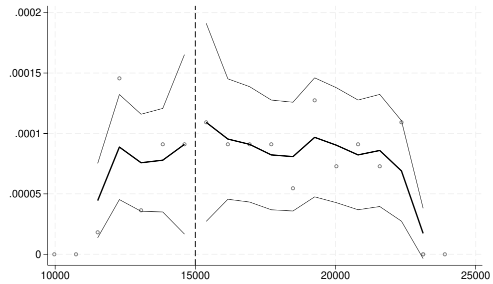

(a) 2,000 population cutoff



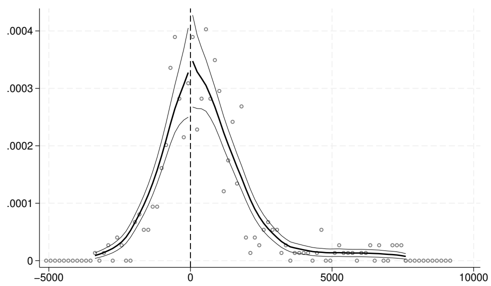

(b) 4,000 population cutoff




(c) 8,000 population cutoff



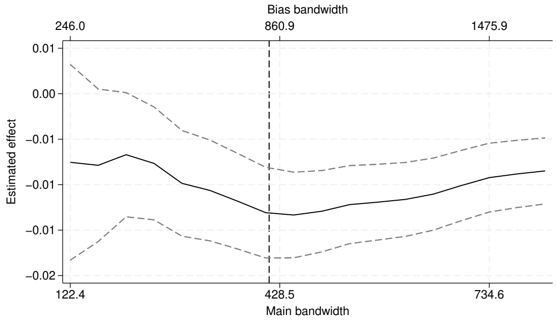

(d) 15,000 population cutoff



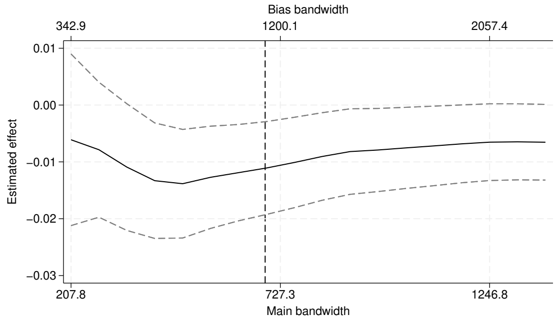

(e) Pooled population cutoff

Figure A9: McCrary density tests.

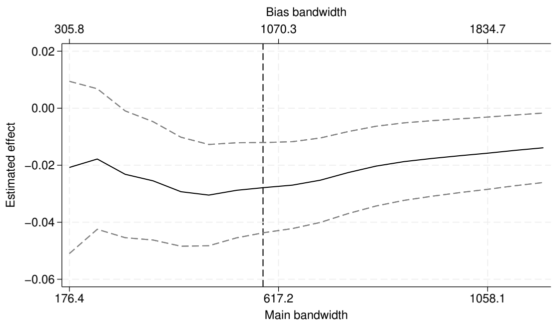
Fourth, in Figure A10, we report our analysis of robustness of the nonparametric main results to the **bandwidth choice**. We report both the clustered and non-clustered results for both cohesion indices. The results are robust across a fair range of bandwidths around the optimal ones.



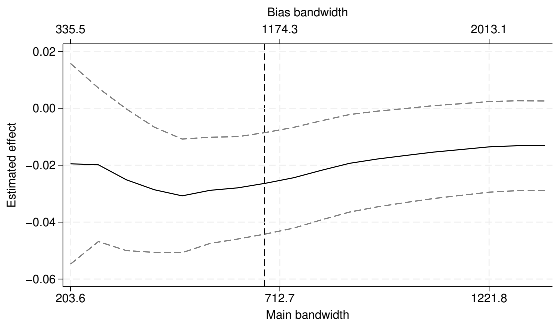

(a) All questions, classical inference




(b) All questions, clustered inference



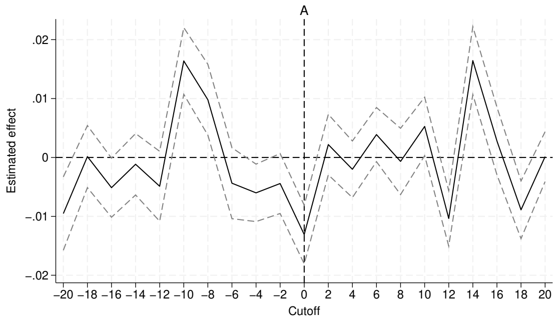

(c) Redistribution questions, classical inference



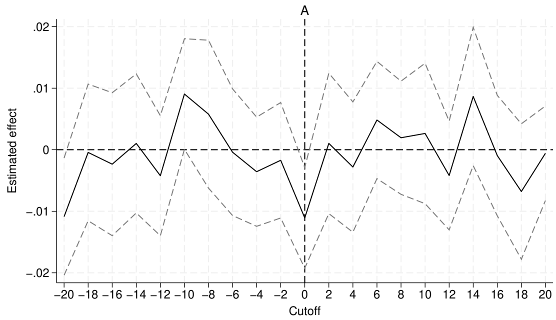

(d) Redistribution questions, clustered inference

Figure A10: Robustness of the results to alternative bandwidths.

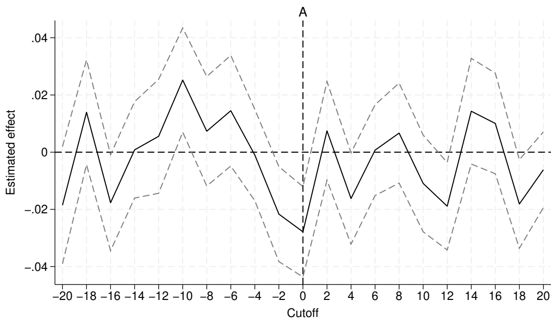
Fifth, in Figure A11, we conduct the **placebo cutoff analysis**. Here, we artificially move the cutoffs away from their real location. The $x$-axis shows by how many percentages we move them away from the original location. Each cutoff is moved by the same relative amount to the same direction at the same time. The real estimate is located at zero in the $x$-axis. The $y$-axis reports the bias-corrected coefficient and the respective robust 95% confidence interval. If the design is valid and the specification appropriate, we should observe that the placebo coefficients are not statistically different from zero (at least not more often than what would arise randomly with multiple testing). This analysis is especially useful for understanding whether the applied RDD specification is appropriate (Hyytinen et al. (2018b)).



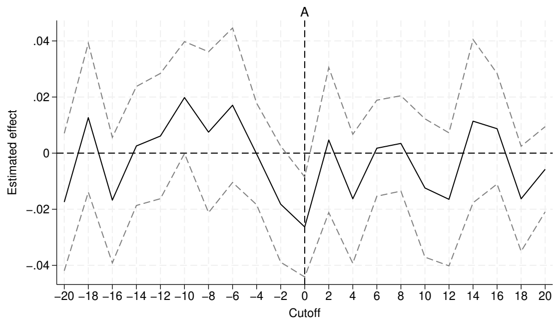

(a) All questions, classical inference




(b) All questions, clustered inference




(c) Redistribution questions, classical inference




(d) Redistribution questions, clustered inference

Figure A11: Robustness of the results to placebo cutoffs.

We observe that the non-clustered results show many significant positive and negative coefficients especially for the redistribution index. This analysis reveals that we should rely on the conventional (non-clustered) results less. This is because there is within-municipality correlation in the policy positions of the candidates. If the bandwidth calculation does not account for this clustering problem, the optimal bandwidths are too narrow in the sense that the results are derived using too few clusters, so, in practice, the placebo effects mainly reflect a small number of municipal fixed effects. This leads to the standard problem that, in small samples, any result is possible by chance even if the design is as-good-as random.

The placebo cutoff analysis for the clustered specification works as it is supposed to as we have non-significant placebo results. There is one exception but that is natural due to multiple testing. Therefore, we feel confident in trusting the clustered results.

Sixth, in Table A2, we present further **balance tests** on additional variables complementary to the ones presented in the main text. We report only the clustered nonparametric results (the non-clustered ones are similar). For further brevity, we only report the bias-corrected point estimates and robust standard errors. We focus here attention on the outcomes that are realized before the elections and, thus, are pre-treatment covariate balance tests. Lyytikäinen and Tukiainen (2019) analyze the effect of the council size on turnout and pivotal probabilities with larger data (more election years and more cutoffs). We do not study these outcomes as they are post-election outcomes that can be influenced by cohesion but not the other way around due to timing. Moreover, any impacts that council size may have on policy outcomes (Pettersson-Lidbom 2012) during the council term are not a concern to us as they, if any, are also realized long after the elections.

Table A2: Covariate balance tests, CTT, clustered

| | **Panel A: Economic and population characteristics of the municipalities** | | | | | | |
|---|---|---|---|---|---|---|---|
| Outcome | Personnel | Tax revenue | Over 65yo | Grants | Expen. | Unemp. | Council Size |
| RD coefficient | -341 | -0.07 | 0.006 | -0.26 | -461 | -1.27 | 6.24** |
| s.e. | 309 | 0.10 | 0.016 | 0.28 | 411 | 1.43 | 2.02 |
| $N$ within main bandwidth | 217 | 207 | 314 | 323 | 286 | 245 | 237 |
| | **Panel B: Candidate level characteristics** | | | | | | |
| Outcome | Unemployed | University | Male | Old | Incumbent | | |
| Avg. effect | 0.004 | -0.007 | -0.013 | -0.040 | -0.060 | | |
| s.e. | 0.012 | 0.029 | 0.030 | 0.031 | 0.036 | | |
| $N$ within main bandwidth | 7290 | 4615 | 6867 | 7349 | 7154 | | |

Notes: All outcome variables are measured at the pre-treatment stage, that is, economic and demographic characteristic are measured before the council term. Results are generated using rdrobust package in STATA (Calonico et al. 2016) using MSE-optimal bandwidth (optimized for each outcome separately) and triangular kernel. We use local linear regression to construct the point estimator. We use local quadratic regression to construct the bias correction. We report the bias corrected and robust point estimates and inference. Standard errors (and bandwidths) are clustered at the municipality level. ** denotes statistical significance at 1% level and * denotes statistical significance at 5% level.

For the balance test purposes, we hope to see non-significant coefficients for other outcomes except for the council size itself (*Council Size*), which should jump at the threshold in order to provide us with enough power for the design and the specifications. That is indeed the case.

For other municipal economic and demographic characteristics, we report municipal personnel per thousand inhabitants (*Personnel*), municipal income tax revenue (*Tax revenue*), share of citizens over the age of 65 (*Over 65yo*), central government transfers in 1,000€ per capita (*Grants*), expenditure in € per capita (*Expen.*), and unemployment rate (*Unemp.*). All these variables are measured in the election year and, thus, cannot yet be influenced by the council that is elected in that year but starts its term in the beginning of the following year. For individual candidate characteristics, we report their unemployment status, a dummy for university education, a dummy for being male, a dummy for being 65 years old or older, and their incumbency status.

The sole unbalanced variable is the individual-level incumbency status which is statistically significant only at the 10% level. It is surprising that this coefficient is negative because, by construct, there are more incumbents in larger councils. As the negative coefficient is hard to rationalize theoretically, it is likely to be just a statistical fluke related to multiple testing.

Table A3: Multiple cutoffs

| | pop.>2k | pop.>4k | pop.>8k | pop.>15k |
|---|---|---|---|---|
| Modified Gallagher | -0.01 | -0.02** | -0.02** | -0.02 |
| Slope index | 0.03 | -0.10** | -0.06* | -0.08 |
| All questions index | -0.008 | -0.006 | -0.018** | 0.016 |
| Redistribution index | -0.004 | -0.035 | -0.029 | 0.06 |
| Incumbency | -0.12 | -0.20 | 0.02 | -0.07 |
| Candidates (munic.) | -1.57 | 1.75 | 3.76 | -6.09 |
| Numb parties | -0.75 | -0.30 | 0.44 | -1.8 |
| Eff numb par | -0.72 | 0.15 | 0.005 | -1.1* |
| Candidates per seat | -0.37 | -0.52 | -0.85 | -0.73 |
| Candidates (party) | 1.22 | 1.25 | 4.63 | -0.31 |
| Respondents (party) | -1.90 | -0.71 | 2.50 | -0.80 |

Notes: Candidates (party) and Respondents (party) are at the party-municipality -election year -level while the other variables are municipality-election year -level. Results are generated using rdrobust package in STATA (Calonico et al. 2016) using MSE-optimal bandwidth (optimized for each outcome and cutoff separately) and triangular kernel. We use local linear regression to construct the point estimator. We use local quadratic regression to construct the bias correction. We report the bias corrected and robust point estimates and inference. Standard errors (and bandwidths) are clustered at the municipality level. ** denotes statistical significance at 1% level and * denotes statistical significance at 5% level.

However, to argue further that the incumbency status does not confound the results, we study in Table A3 whether incumbency changes most at those cutoffs where distortions respond the most. This is not the case. Possible alternative mechanisms could be that the number of candidates, of parties or of respondents changes at the cutoffs at which distortions change. However, we show in the same table that their coefficients are not statistically significant, and cohesion measures are the only variables that show a significant effect at the same cutoff as the distortions. Thus, the overall evidence gives the strongest support to the distortions being behind the cohesion response.

Table A4: Do candidates respond differently if their contextual environment changes?

| | Daycare | Social welfare | Anti environment | Progressive | Sell | Subsidy |
|---|---|---|---|---|---|---|
| Party switcher | 0.0491 | -0.0483 | 0.0306 | 0.1414** | 0.0097 | -0.0308 |
| s.e. | [0.0754] | [0.0504] | [0.0567] | [0.0611] | [0.0252] | [0.0245] |
| Constant | 1.5462** | 1.1781** | 0.5947** | 1.0411** | 0.4801** | 0.1580** |
| s.e. | [0.0920] | [0.0552] | [0.0739] | [0.0927] | [0.0319] | [0.0345] |
| N | 7504 | 7504 | 7504 | 7504 | 7504 | 7504 |
| | Daycare | Social welfare | Anti environment | Progressive | Sell | Subsidy |
| Absolute change in council size | 0.0009 | 0.0005 | 0.0016 | 0.0000 | 0.0002 | -0.0002 |
| s.e. | [0.0015] | [0.0011] | [0.0019] | [0.0015] | [0.0009] | [0.0008] |
| Constant | 1.3847** | 1.0241** | 0.8555** | 0.9922** | 0.3356** | 0.1017** |
| s.e. | [0.0952] | [0.0540] | [0.0685] | [0.0815] | [0.0307] | [0.0383] |
| N | 7504 | 7504 | 7504 | 7504 | 7504 | 7504 |

Notes: Outcomes are absolute changes in responses to following questions. *Daycare*: "If one of the parents is at home, we should limit the right of the family to have their child placed in daycare". *Social welfare*: "It is nowadays too easy to be admitted to social welfare". *Anti environment*: "We can relax environmental policies if it increases employment or material welfare of municipal residents.". *Progressive*: "The municipal user fees should be made more progressive in income". *Sell*: "Which of the following options should be mainly used in order to balance the municipal budget in your municipality? Choose two of the following options. Option: Selling off municipal property (2012 and 2008 are similar but not identical)". *Subsidy*: "Which of the following options should be mainly used in order to balance the municipal budget in your municipality? Choose two of the following options. Option: Developing the business in the municipality (2012 and 2008 are similar but not identical)". Party switcher is a dummy for candidates running in different parties in 2008 and 2012. Absolute change in council size is a continuous variable measuring how much council size changed in absolute terms between 2012 and 2008. All regressions control for municipal and party label fixed effects, incumbency status, university education, unemployed status, student status, high professional sose status, gender, being above 60 years old, and age. Sample includes 2012 candidates that ran and responded to the VAA both in 2008 and 2012. Standard errors clustered at municipality level reported in brackets. ** denotes statistical significance at 1% level and * denotes statistical significance at 5% level.

Seventh, in Table A4 we provide evidence that candidates do not change their policy position in VAA when their incentives change. To achieve this, we analyze within candidate changes between 2008 and 2012 in those VAA questions that remain the same or very similar across the years, when candidates' contextual environment changes. First, we analyze changes in council size, which may happen due to changes in municipal population, municipal mergers or candidate mobility. Whichever the reason, all should affect candidates' incentives to respond strategically. Second, we analyze party switching. Out of 8550 candidates in total that re-run and respond to survey both years 5.3% switch their party. We study absolute changes so that possible changes in the opposite directions do not cancel each other out in the estimation. While candidates overall change their responses somewhat over time (*Constant*), they do not respond to changes in their contextual environment. Only one out of 12 estimates of interest in Table A4 is statistically significant and the magnitudes of the point estimates in relation to the constants are small. These results support the view that is not the candidates who respond to changes in council size but rather the plausible mechanism is candidate selection by parties.

Eighth, in Table A5, we replicate our main result when **including all municipalities**, that is, not limiting the sample to municipalities with population smaller than 22500. The point estimates are smaller than in the main results. This is expected theoretically as demonstrated in Figure A4, because at larger population cutoffs the effects should be smaller. The effective sample is much larger than in the main analysis both because of the larger bandwidths and including larger municipalities.

Table A5: All cutoffs included: Policy positions and council size, nonparametric RDD (candidate level)

| | **All questions** | **All questions** | **Redistribution** | **Redistribution** |
|---|---|---|---|---|
| Conventional local linear RD coefficient | -0.0039 | -0.0038 | -0.0096 | -0.0091 |
| 95% Confidence interval with bias-correction and robust inference | [-0.006 ; -0.002] | [-0.009 ; 0.0001] | [-0.017 ; -0.004] | [-0.019 ; -0.002] |
| N within main bandwidth | 16130 | 16211 | 16503 | 16529 |
| MSE-optimal bandwidths (main/bias) | 3450/8871 | 3511/8011 | 3141/8004 | 3322/8196 |
| Clustered bandwidths and s.e.'s | No | Yes | No | Yes |

Notes: Results are generated using rdrobust package in STATA (Calonico et al. 2016) using MSE-optimal bandwidth and triangular kernel. We use local linear regression to construct the point estimator. We use local quadratic regression to construct the bias correction. Sample is not limited to municipalities with population below 22500.

Table A6: Elected candidates only: Policy positions and council size, nonparametric RDD (candidate level)

| | **All questions** | **All questions** | **Redistribution** | **Redistribution** |
|---|---|---|---|---|
| Conventional local linear RD coefficient | -0.007 | -0.0064 | -0.021 | -0.020 |
| 95% Confidence interval with bias-correction and robust inference | [-0.014 ; -0.0029] | [-0.015 ; -0.0008] | [-0.049 ; -0.004] | [-0.051 ; -0.002] |
| N within main bandwidth | 1828 | 2128 | 1758 | 1907 |
| MSE-optimal bandwidths (main/bias) | 643/1128 | 717/1191 | 593/1049 | 641/1077 |
| Clustered bandwidths and s.e.'s | No | Yes | No | Yes |

Notes: Results are generated using rdrobust package in STATA (Calonico et al. 2016) using MSE-optimal bandwidth and triangular kernel. We use local linear regression to construct the point estimator. We use local quadratic regression to construct the bias correction. The sample includes only the elected candidates.

Ninth, in Table A6, we replicate our main result when **restricting our sample to the elected candidates**. As our results indicate, elected councils consist of more heterogeneous parties as municipalities grows larger. That is, our main result of council size influencing parties' selection incentives when recruiting candidates for their lists seems to carry over to the elected governing body.

Tenth, we report 2SLS results in Tables A7 and A8. We show the second stage, first stage and the corresponding OLS results. All the coefficients across all specifications have the expected sign in the second stage (and the first stage and the OLS), that is, as the realized distortions increase party heterogeneity increases. However, the second stage results are not statistically significant at the 5% level (lowest p-value 0.095), and while the first-stage relationship is often significant, the F-test is in 6 out of 8 specifications below the classic rule of thumb 10. This is expected, because the 2SLS estimation is always low-powered compared to a reduced form IV specification, where our RDD on cohesion outcomes can be seen as the latter. Moreover, compared to the reported RDD results regarding the distortion indices in Table 1 of the paper (the first stage), we lose part of the sample in the 2SLS due to the availability of enough VAA responses for the second stage analysis of effects on cohesion, further reducing power in the 2SLS first stage. Finally, the second stage coefficients are much larger than the OLS ones. However, given that the second stage estimates are very noisy, the difference between them and the OLS estimates is not statistically significant, and thus, this difference in magnitude does not offer much information.

Table A7: 2SLS estimation results for the effect of modified Gallagher index on the policy position indices

| | **All questions** | | **Redistribution** | |
|---|---|---|---|---|
| Second stage estimate | 0.258 | 0.401 | 0.641 | 2.679 |
| Standard error | 0.154 | 0.414 | 0.428 | 1.744 |
| First stage estimate | -0.0105** | -0.0083* | -0.105** | -0.0089 |
| Standard error | 0.0027 | 0.0040 | 0.0026 | 0.0048 |
| OLS estimate | 0.087** | 0.071 | 0.207** | 0.256 |
| Standard error | 0.034 | 0.051 | 0.101 | 0.179 |
| N within bandwidth | 474 | 238 | 474 | 159 |
| Bandwidth | All data | MSE-optimal | All data | MSE-optimal |
| First stage F-test | 16.18 | 4.24 | 16.18 | 3.47 |
| Outcome mean (std. dev.) | 0.136 (0.013) | 0.136 (0.014) | 0.292 (0.032) | 0.291 (0.039) |

*Notes: Unit of observation is a municipality-year. Instrumental variable is a dummy for being above a council size cutoff. Linear trend of the forcing variable on both sides of the cutoff is controlled for. Columns (2) and (4) limit the sample within the MSE optimal bandwidths for the main outcome RDD. Standard errors are clustered at the municipality level. * p<0.05, ** p<0.01.*

Table A8: 2SLS estimation results for the effect of slope index on the policy position indices

| | **All questions** | | **Redistribution** | |
|---|---|---|---|---|
| Second stage estimate | 0.149 | 0.204 | 0.370 | 1.549 |
| Standard error | 0.107 | 0.274 | 0.284 | 1.829 |
| First stage estimate | -0.0181* | -0.0163 | -0.0181* | -0.0155 |
| Standard error | 0.0088 | 0.0147 | 0.0088 | 0.0174 |
| OLS estimate | 0.018* | 0.015 | 0.065** | 0.088** |
| Standard error | 0.011 | 0.015 | 0.025 | 0.044 |
| N within bandwidth | 474 | 238 | 474 | 159 |
| Bandwidth | All data | MSE-optimal | All data | MSE-optimal |
| First stage F-test | 4.26 | 1.22 | 4.26 | 0.79 |
| Outcome mean (std. dev.) | 0.136 (0.013) | 0.136 (0.014) | 0.292 (0.032) | 0.291 (0.039) |

*Notes: Unit of observation is a municipality-year. Instrumental variable is a dummy for being above a council size cutoff. Linear trend of the forcing variable on both sides of the cutoff is controlled for. Columns (2) and (4) limit the sample within the MSE optimal bandwidths for the main outcome RDD. Standard errors are clustered at the municipality level. * p<0.05, ** p<0.01.*

# A3 Descriptive statistics

Table A9: Respondent sample selection

| | **Non-respondents** | | | **Respondents** | | |
|---|---|---|---|---|---|---|
| | n | Mean | Std.Dev. | n | Mean | Std.Dev. |
| Male | 34,795 | 0.63 | 0.48 | 40,839 | 0.58 | 0.49 |
| Age | 34,794 | 50.9 | 13.5 | 40,839 | 45.9 | 13.0 |
| Incumbent | 34,795 | 0.18 | 0.38 | 40,839 | 0.23 | 0.42 |
| Votes | 34,794 | 39 | 61 | 40,839 | 90 | 204 |
| Elected | 34,795 | 0.21 | 0.40 | 40,839 | 0.32 | 0.46 |
| University | 27,325 | 0.11 | 0.31 | 37,040 | 0.24 | 0.43 |
| Unemployed | 34,772 | 0.06 | 0.24 | 40,819 | 0.05 | 0.21 |
| Old | 34,795 | 0.28 | 0.45 | 40,839 | 0.15 | 0.36 |

Notes: Comparison of 2008 and 2012 Finnish municipal candidates who responded or did not respond to the YLE VAA survey.

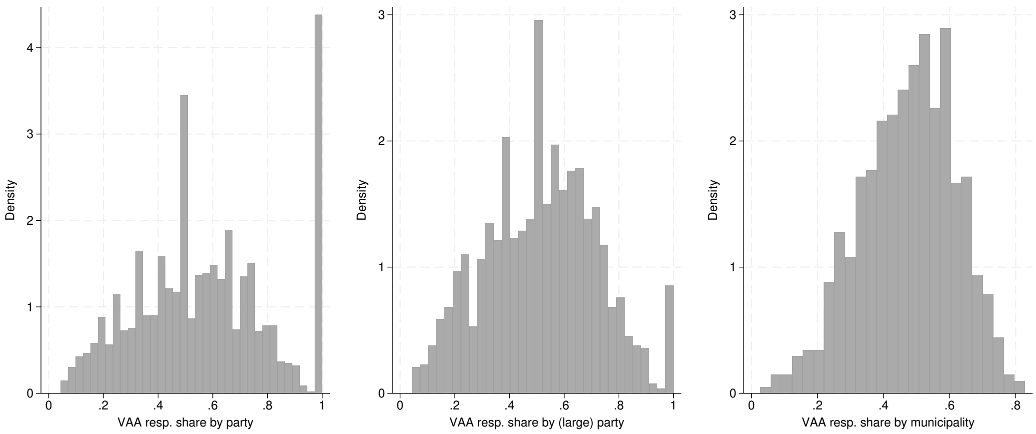



Figure A12: Histograms of VAA response shares by party or municipality. On the left, by party for all parties, in the middle, by party only for the three largest parties nationally (KESK, KOK, SDP), and on the right, by municipality.

Table A10: Descriptive statistics

**Panel A: Economic and population characteristics of the municipalities**

| | Personnel | Tax revenue | Over 65yo | Grants | Expen | Unemp | Population |
|---|---|---|---|---|---|---|---|
| mean | 2271 | 2.82 | 0.22 | 2.26 | 6300 | 11.9 | 6846 |
| s.d. | 774 | 0.47 | 0.05 | 0.81 | 1094 | 3.71 | 5143 |

**Panel B: Political characteristics of the municipalities**

| | Council size | Candidates | Numb parties | Eff numb par | Redistribution | All questions |
|---|---|---|---|---|---|---|
| mean | 27.6 | 81 | 5.91 | 3.49 | 0.29 | 0.14 |
| s.d. | 7.70 | 38.5 | 1.44 | 0.92 | 0.03 | 0.01 |

**Panel C: Party level characteristics**

| | Vote share | Respondents | Candidates | Cand per seat | Redistribution | All questions |
|---|---|---|---|---|---|---|
| mean | 0.28 | 12.7 | 23.5 | 3.10 | 0.29 | 0.14 |
| s.d. | 0.16 | 6.5 | 11.3 | 0.73 | 0.04 | 0.02 |

**Panel D: Candidate level characteristics**

| | Unemployed | University | Male | Old | Incumbent | Redistribution | All questions |
|---|---|---|---|---|---|---|---|
| mean | 0.04 | 0.19 | 0.58 | 0.15 | 0.28 | 0.29 | 0.14 |
| s.d. | 0.19 | 0.40 | 0.49 | 0.36 | 0.45 | 0.10 | 0.03 |

## A4 Voting aid application questions

### YLE voting aid application questions in 2008

- In order to provide our municipality with more revenue, we should [choose two]:
  - increase the property tax rate for residential buildings. (Redistribution index)
  - increase the property tax rate for holiday houses. (Redistribution index)
  - increase user fees. (Redistribution index)
  - introduce new user fees. (Redistribution index)
  - sell off municipal property.
  - consider a municipality merger.
  - attract business with favorable conditions or financial support.
  - attract new well-off taxpayers by offering them building plots.
  - request for more state subsidies.
- Which of the following services should we privatize [choose as many as you like but at least one of the following]:
  - comprehensive school.
  - health center.
  - eldercare.
  - day care.
  - municipal engineering.
  - social welfare.
  - substance abuse treatment and rehabilitation.
  - fire and rescue services.
  - zoning.
  - special health care.
  - water utility.
  - none of the above.
- The following questions have a five-step scaling: 0 = completely disagree, 1 =

somewhat disagree, 2 = empty, 3 = somewhat agree, 4 = completely agree

  - It is nowadays too easy to be admitted to social welfare. (Redistribution index)
  - The municipal user fees should be made more progressive in income. (Redistribution index)
  - If there is no other option, we should raise the municipal tax rate rather than cut from the municipal services.
  - If one of the parents is at home, we should limit the right of the family to have their child placed in daycare.
  - We should downsize the number of employees in my municipality because there are too many of them.
  - We can relax environmental policies if it increases employment or material welfare of municipal residents.
  - If forced to choose, it is better to cut from healthcare than from schools because there are private health centers but no private schools.

**YLE voting aid application questions in 2012**

- Which of the following options should be mainly used in order to balance the municipal budget in your municipality? Choose two of the following options:
  - issuing more debt. (Redistribution index)
  - increasing user fees or introduction of new ones. (Redistribution index)
  - raising taxes. (Redistribution index)
  - cutting down services.
  - selling off municipal property.
  - developing the business in the municipality.
- If my municipality got a big donation to be spent on municipal services, it should mainly spent on [choose two]:
  - social welfare
  - daycare
  - eldercare

  - schools
  - wages of nurses
  - health centers
  - special healthcare

- Let's assume that your municipality is financially troubled. You must save and there is a trade-off between the services for the elderly and the children. What will you do?
  - We should save but I still propose issuing more debt. (Redistribution index)
  - I cut from the services for the elderly.
  - I cut from the services for the children.
  - I try to cut even-handedly from both kinds of services.

- A historical building has fallen into disrepair and it has been proposed that it should be demolished. On the other hand, there are proposals that it should be repaired. What is your stance?
  - History is more important than a temporary economic gain. The building should be repaired.
  - Economic gain is more valuable than history.
  - The plot should be used more gainfully than now.
  - There is a risk of a precipitate decision. I suggest we postpone the decision.
  - I listen at what municipal residents opine on the matter and form my opinion accordingly.
  - Historical value and economic gain are not comparable. I don't want to take a stand.

- They're planning a new residential area where immigrants and people living on social welfare would live next to well-off. How do you find it?
  - I definitely disapprove. I can't comprehend such an attempt towards a diverse area.
  - I try to consider if the project could create something socially and humanely

new that we could learn from.

  - I actively start looking for a compromise without taking a firm stand.
  - I find the project very positive. I'm just amazed by intolerance of the others.

- The following questions have a five-step scaling: 0 = completely disagree, 1 = somewhat disagree, 2 = empty, 3 = somewhat agree, 4 = completely agree
  - We should increase the health care user fees in my municipality. (Redistribution index)
  - It is nowadays too easy to be admitted to social welfare. (Redistribution index)
  - We should raise the property tax rate in my municipality. (Redistribution index)
  - The municipal user fees should be made more progressive in income. (Redistribution index)
  - The old should have a universal right to a retirement home similar to one enjoyed now by children and daycare.
  - Privatization of municipal health care would increase efficiency and lower the costs.
  - If one of the parents is at home, we should limit the right of the family to have their child placed in daycare.
  - My municipality should receive refugees with a positive asylum decision.
  - Too little attention has been paid to marginalization of children and youth in my municipality.
  - Public bins of my municipality should allow for recycling.
  - More public money should be spent on road maintenance.
  - We can relax environmental policies if it increases employment of municipal residents.
  - If my municipality were to merge with another municipality, there should be a non-binding referendum about the merger.
  - The five-year long dismissal period for the municipal employees in conjunction

with a municipality merger is too long.

- Municipal employees should not be nominated as municipal board members.
- The voting age in municipal elections should be decreased to 16 years [from 18 years].
- MPs should not run in municipal elections.

# References

Bol, D., K. Matakos, O. Troumpounis, and D. Xefteris (2019). Electoral rules, strategic entry and polarization. *Journal of Public Economics 178*, 104065.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2016). rdrobust: Software for regression discontinuity designs. *Stata Journal 17*(2), 372–404.

Dal Bò, E. and F. Finan (2018). Progress and perspectives in the study of political selection. *Annual Review of Economics 10*(1), 541–575.

Debreu, G. (1952). A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences 38*(10), 886–893.

Duverger, M. (1954). *Political parties: Their organization and activity in the modern state*. Methuen.

Flis, J., W. Słomczyński, and D. Stolicki (2020). Pot and ladle: A formula for estimating the distribution of seats under the Jefferson–D'Hondt method. *Public Choice 182*(1-2), 201–227.

Gallagher, M. (1991). Proportionality, disproportionality and electoral systems. *Electoral Studies 10*(1), 33–51.

Harjunen, O., T. Saarimaa, and J. Tukiainen (2023). Love thy (elected) neighbor? Residential segregation, political representation and local public goods. *Journal of Politics 85*(3).

Herrera, H., A. Llorente-Saguer, and J. C. McMurray (2019). The marginal voter's curse. *Economic Journal 129*(624), 3137–3153.

Herrera, H., M. Morelli, and S. Nunnari (2016). Turnout across democracies. *American Journal of Political Science 60*(03), 607–624.

Herrera, H., M. Morelli, and T. R. Palfrey (2014). Turnout and power sharing. *Economic Journal 124*(574), F131–F162.

Herron, E. S., R. J. Pekkanen, and M. S. Shugart (2018). *The Oxford handbook of electoral systems*. Oxford University Press.

Hyytinen, A., J. Meriläinen, T. Saarimaa, O. Toivanen, and J. Tukiainen (2018a). Public employees as politicians: Evidence from close elections. *American Political Science Review 112*(1), 68–81.

Hyytinen, A., J. Meriläinen, T. Saarimaa, O. Toivanen, and J. Tukiainen (2018b). When does regression discontinuity design work? Evidence from random election outcomes. *Quantitative Economics 9*, 1019–1051.

Iaryczower, M. and A. Mattozzi (2013). On the nature of competition in alternative electoral systems. *Journal of Politics 71*(3), 829–853.

Lyytikäinen, T. and J. Tukiainen (2019). Are voters rational? *European Journal of Political Economy 59*, 230–242.

Matakos, K., O. Troumpounis, and D. Xefteris (2016). Electoral rule disproportionality and platform polarization. *American Journal of Political Science 60*(4), 1026–1043.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Meriläinen, J. (2022). Political selection and economic policy. *The Economic Journal 132*(648), 3020–3046.

Osborne, M. J. and A. Slivinski (1996). A model of political competition with citizen-candidates. *Quarterly Journal of Economics 111*(1), 65–96.

Pettersson-Lidbom, P. (2012). Does the size of the legislature affect the size of government? Evidence from two natural experiments. *Journal of Public Economics 96*(3), 269–278.

Romer, T. and H. Rosenthal (1978). Political resource allocation, controlled agendas, and the status quo. *Public Choice 33*(4), 27–43.

Romer, T. and H. Rosenthal (1979). Bureaucrats versus voters: On the political economy of resource allocation by direct democracy. *Quarterly Journal of Economics 93*(4), 563–587.

Schuster, K., F. Pukelsheim, M. Drton, and N. R. Draper (2003). Seat biases of apportionment methods for proportional representation. *Electoral Studies 22*(4), 651–676.

Taagepera, R. (1986). Reformulating the cube law for proportional representation elections. *American Political Science Review 80*(2), 489–504.

Theil, H. (1969). The desired political entropy. *American Political Science Review 63*(2), 521–525.