

Electoral Institutions and Intraparty Cohesion

Konstantinos Matakos

Riikka Savolainen

Orestis Troumpounis

Janne Tukiainen

Dimitrios Xefferis

VATT WORKING PAPERS

109

Electoral Institutions and Intraparty Cohesion

Konstantinos Matakos
Riikka Savolainen
Orestis Troumbounis
Janne Tukiainen
Dimitrios Xefteris

Konstantinos Matakos, Corresponding author: Department of Political Economy, King's College London, Strand campus, London, WC2R 2LS, UK; email: konstantinos.matakos@kcl.ac.uk

Riikka Savolainen, Department of Political Economy, King's College London, Strand campus, London, WC2R 2LS, UK.

Orestis Troumpounis, Department of Economics, University of Padova, via del Santo 33, 35123 Padova, Italy and Department of Economics, Lancaster University, Bailrigg, Lancaster, LA1 4YX, United Kingdom.

Janne Tukiainen, Department of Government, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, United Kingdom, and VATT Institute for Economic Research, P.O. Box 1279 (Arkadiankatu 7), FI-00101 Helsinki, Finland.

Dimitrios Xefteris, Department of Economics, University of Cyprus, PO Box 20537, 1678 Nicosia, Cyprus.

We are grateful to James Adams, Manuel Bagues, Peter Buisseret, Alessandra Casella, John Duggan, Olle Folke, Alexander Fourinaies, Anthony Fowler, Bernard Grofman, Tasos Kalandrakis, David Kang, Eva Mörk, Tuomas Pekkarinen, Carlo Prato, Johanna Rickne, Jim Snyder, Stephane Wolton 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 Chicago Harris, Columbia, Harvard, HECER, King's College London, Lancaster, Leicester, LSE, VATT and Zurich.

This version (March 4, 2019) replaces the July 12, 2018 version.

ISBN 978-952-274-222-3 (PDF)

ISSN 1798-0291 (PDF)

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Arkadiankatu 7, 00100 Helsinki, Finland

Helsinki, March 2019

Electoral Institutions and Intraparty Cohesion

VATT Institute for Economic Research
VATT Working Papers 109/2018

Konstantinos Matakos – Riikka Savolainen – Orestis Troumbounis – Janne Tukiainen – Dimitrios Xefteris

Abstract

We study parties' optimal ideological cohesion across electoral rules, when the following trade-off is present: A more heterogeneous set of candidates is electorally appealing (catch-all party), yet, it serves policy-related goals less efficiently. When the rule becomes more disproportional, thus inducing a more favorable seat allocation for the winner, the first effect is amplified, incentivizing parties to be less cohesive. We provide empirical support using a unique data-set that records candidates' ideological positions in Finnish municipal elections. Exploiting an exogenous change of electoral rule disproportionality at different population thresholds, we identify the causal effect of electoral rules on parties' cohesion.

Key words: keyword, keyword, keyword, keyword, keyword

JEL classes: C21; C72; D02; D72

1 Introduction

The internal structure of an organization can be an important determinant of its performance in competitive environments (Grossman and Hart 1986; Aghion and Tirole 1997; Besley and Persson 2018; Beal et al. 2003). For example, an organization’s cohesion may affect knowledge transfers (Morris 2000; Reagans and McEvily 2003; Acemoglu et al. 2011), and increase workers’ marginal products (Levine 1991). In the political arena, cohesive parties tend to vote in a disciplined manner and therefore guarantee the survival of governments and effective policy implementation (Carey 2008; Tsebelis 2002). On the other hand, diversified organizations can facilitate the diffusion of certain technologies (e.g., Reich 2016), and have the ability to address a larger pool of consumers or voters (e.g., Borenstein 1991; Kirchheimer 1990). Therefore, the optimal degree of organizational cohesion is not a trivial decision and, certainly, not one that is typically considered in an institutional vacuum. Indeed, the following question arises: How does the underlying institutional framework determine organizations’ internal structure in the first place?

In this paper, we study how institutions affect organizations’ structure via the channel of recruitment. In particular, we focus on political parties and explore, both theoretically and empirically, how electoral institutions shape parties’ strategic incentives at the stage of candidate selection.¹ As Dal Bò and Finan (2018) conclude in their survey, existing literature on candidate selection (see e.g. Besley et al. 2017; Dal Bò et al. 2017; Folke and Rickne 2017; Beath et al. 2016) leaves unexplored a potentially important factor: the *strategic role political parties play in candidate selection*. Indeed, in some cases (e.g. in the U.S.), political parties have limited control over candidate nominations, but, in many others (e.g. in Europe and in most parliamentary systems), parties have almost absolute power in candidate selection. How do different electoral institutions affect parties’ incentives regarding which candidates to nominate for office? What determines whether a party nominates a more or less homogeneous –in terms of policy preferences– list of candidates? In this paper we try to provide some first insights by conducting a formal analysis and by testing its results in a quasi-experimental framework.²

¹Electoral systems, a key institutional determinant of political competition, are known to affect parties’ incentives and, hence, a variety of political and economic outcomes, for example polarization (Cox 1990; Matakos et al. 2016), 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 (Persson et al. 2003; Milesi-Ferretti et al. 2002; Persson et al. 2007; Lizzeri and Persico 2001). The literature on electoral systems is vast and we just refer to some representative examples among many others. See for more references therein as well as Persson and Tabellini (2002, 2005); Lijphart (1995, 1999); Taagepera and Shugart (1989); Grofman (2008).

²Diermeier and Feddersen (1998) focus on legislative cohesion due to the possibility of vote of confidence. Cirone (2017) explores the role of dual mandates on voting cohesion and early party institutionalization, Buisseret and Prato (2017) study the effects of electoral rules on parties’ legislative cohesion -yet

We refer to the electoral institutions in an easy-fitting manner by focusing on the disproportionality of the electoral system. As well known, electoral rules that are proportional map electoral results to parliamentary power in an accurate manner. On the contrary, disproportional electoral rules favor the election winner (via various specific characteristics, e.g., the electoral formula, the district magnitude, the presence of thresholds for representation or the size of the body to be elected (Herron et al. 2018; Lijphart 1995)). In our “reduced-form” approach, we simply associate disproportional electoral rules to a favorable distribution of parliamentary power for the election winner. Our results hence have a broad appeal, since they apply across different institutional settings, and do not hinge on particular institutional parameters.

Our analysis of parties’ ideological cohesion refers to the ideological homogeneity of their candidates. We think of parties as a collection of candidates *“tethered by a rubber band to the ideology espoused by the parties whose label they run on”* (Grofman 2008). By linking the electoral rule disproportionality to parties’ ideological cohesion we are able to determine both theoretically and empirically how elastic this rubber band is, and, thus, how much candidates’ ideologies deviate from parties’ positions under different levels of disproportionality. Our results suggest that proportional rules are associated with ideologically cohesive parties, while disproportional ones lead to pluralistic parties that embrace a rather heterogeneous set of candidates.

Our main message is that electoral institutions affect the incentives that political organizations face in recruiting their political personnel due to the following key trade-off: wider ideological variety amongst candidates implies more votes. At the same time, it also implies higher costs and lower efficiency -for instance, less cohesive parties are more likely to produce policy outcomes that deviate from those desired by the leadership. Therefore, the gain-loss tradeoff associated to a marginal increase in vote share, determines the optimal variety, and, thus, the electoral rule critically drives parties’ behavior. Depending on how the electoral rule transforms votes into seats and political influence, it affects the importance assigned to votes and, hence, variety. As our theoretical results show, disproportional electoral systems make parties’ incentives to increase their vote share more pronounced, and this is achieved by proposing a relatively wide-ranging set of candidates. On the contrary, electoral rules that are relatively proportional provide less incentives for an increased vote share and are therefore associated with high levels of intraparty ideological bonds.

We find that the above theoretical arguments are empirically supported by identifying the causal effect of the electoral rule disproportionality on parties’ ideological cohesion.

via a mechanism that is different to ours, and Kselman (2017) examines another feature of PR electoral systems (closed vs. open list). Galasso and Nannicini (2017) compare candidates’ quality across electoral rules.

To obtain data on parties' ideological cohesion, we leverage on a unique data-set recording the policy positions of individual candidates for the Finnish municipal elections in years 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.

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. We use changes in the council size as a proxy for changes in the rule's disproportionality allowing for a regression discontinuity design (RDD). As we show in the Appendix, indeed the electoral rule in the municipalities with small-size councils favors the large parties disproportionately (Herron et al. 2018; Benoit 2000). Then, our main results show that, in the elections for those smaller councils, competing lists tend to be less cohesive than in the elections for larger councils.

Finally, we rule out alternative mechanisms that could explain our empirical results. In particular, while one could expect that changes in the council size have other political consequences besides its impact on proportionality (e.g., affecting the number of parties or candidates), we use our rich data to perform extensive covariate balance tests and show that proportionality is the most likely mechanism for the reported effects. We also show that there is no sorting across the thresholds, which is natural as the municipal population is not self-reported. Furthermore, no other policy changes take place at the thresholds that determine councils' sizes. The results are robust to a 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.

Our estimates on the effect of the electoral rule disproportionality (via the council size) on parties' cohesion constitute a novel finding. More generally, the evidence on the causal effects of electoral systems on any outcome is fairly scarce. Typically, researchers have relied on cross-country or panel variation in electoral systems leaving more room to suspect confounding. In contrast, we leverage on plausibly as-good-as-random variation within a country. Moreover, for example, Shugart (2013), and Buisseret and Prato (2017) argue that we have a very limited overall understanding of the mechanisms between the electoral institutions and the behavior of political actors. We start to fill these gaps both theoretically and empirically by proposing and testing a novel mechanism through which election systems influence the parties' candidate selection on ideology.

Overall, our work contributes towards a better understanding of politicians' characteristics in representative democracies.³ While the elected officials' characteristics and

³The 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 A10).

policy positions are known to matter for policy outcomes in several environments (see e.g. Besley et al. 2011; Chattopadhyay and Duflo 2004; Jones and Olken 2005; Lee et al. 2004; Washington 2008) –including the context of our empirical exercise and via an intra-party channel (Hyytinen et al. 2018a; Meriläinen 2018)–, different systems seem to elect politicians with dissimilar traits (Best et al. 2000). Variation in electoral institutions has been offered as one potential explanation (Beath et al. 2016; Carey and Shugart 1995; Gagliarducci and Nannicini 2013; Galasso and Nannicini 2011, 2017) for differences in political selection, and the present paper is, to our knowledge, the first to provide causal evidence in favour of electoral institutions affecting political selection in ideology.

In Section 2 we develop our theoretical arguments, in Section 3 we present our empirical evidence and then, in Section 4, we conclude. All proofs, as well as further discussions of our data, theoretical and empirical results (e.g., robustness) are included in the Appendix.

2 Theory

We present 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.

The policy space is assumed to be continuous, unidimensional, and represented by the interval $X = [0, 1]$. 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 . 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 $x_R \in [1/2, 2/3]$.⁴ Each party j strategically chooses an interval $[\underline{x}_j, \bar{x}_j]$ with $0 \leq \underline{x}_L < \bar{x}_L \leq 1/2$ and $1/2 \leq \underline{x}_R < \bar{x}_R \leq 1$ where its candidates belong to maximize the party’s payoff.⁵

The game evolves as follows: Parties strategically choose their list of candidates and voters vote for the party that included in its list the ideologically closest candidate to them. Given the electoral rule in place, electoral outcomes translate to a distribution of seats in the parliament assigned to candidates of different ideologies. 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 $[\underline{x}_L^*, \bar{x}_L^*]$ and

⁴In the paper we present the simplest version of our model that presents a set of interesting results on intraparty cohesion. After we present our main result, we discuss several ways we could relax some of our assumptions without changing the qualitative features of our equilibrium.

⁵Assuming 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 Appendix (Section A1.1), in the equilibrium, parties would never propose a point on the strategy space.

$[\underline{x}_R^*, \bar{x}_R^*]$ such that $\bar{x}_L^* + \underline{x}_R^* = 1$, $\underline{x}_L^* + \bar{x}_R^* = 1$ and none of the two parties has incentives to propose a different list of candidates.

The electoral system: The electoral system in our model translates each party's vote share v_j into a seat share in the parliament s_j . Let us first describe how v_j is determined given the proposed intervals $[\underline{x}_L, \bar{x}_L]$ and $[\underline{x}_R, \bar{x}_R]$. The indifferent voter is located at $\dot{x} = \frac{\bar{x}_L + \underline{x}_R}{2}$. 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:

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

To capture the electoral institutions, the crucial element is how a party's vote share translates to its seat share. In general, the electoral system is a function $G(v_L, v_R)$ that translates vote shares to seat shares, where $s_L = G(v_L, v_R)$ and $s_R = 1 - G(v_L, v_R)$. Notice now, that given that $v_L + v_R = 1$ seat shares can be written as a function of v_L where $s_L = G(v_L)$ and $s_R = 1 - G(v_L)$. Regarding the properties of the electoral institution $G(v_L)$ we assume that $G(v_L) : [0, 1] \rightarrow [0, 1]$ is continuous, symmetric around 0.5 (i.e., $G(v_L) = 1 - G(1 - v_L)$), $G(0) = 0$, and log concave wherever it takes interior values (i.e., $\frac{\partial^2 \ln G(v_L)}{\partial v_L^2} < 0$ for all v_L such that $G(v_L) \in (0, 1)$).

Examples: One can think of several examples of $G(v_L)$ that could be a part of our analysis. *Theil's rule* (Theil 1969) is a well known method of introducing distortions in favor of the winner across different electoral systems where:

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

and $n \geq 1$ (see Matakos et al. (2016); Herrera et al. (2016, 2019, 2014); Saporiti (2014); Debowicz et al. (2016) for recent applications). If $n = 1$, each party's vote share is equal to its seat share and seats are allocated to parties proportionally to their vote shares. If $n > 1$, the electoral system is allocating disproportionately more seats to the party with the highest vote share. This advantage for the large party increases in n . Figure 1a illustrates the case of a pure PR system ($n = 1$), a relatively disproportional system favoring the first party ($n = 3$, the so called Cube's law considered a "standard" approximation of first-past-the-post systems with several districts), and a hypothetical extreme case where the winner is allocated all the seats in the parliament ($n \rightarrow \infty$).

Alternatively, as Figure 1b illustrates, one could introduce distortions in favor of the winner according to a simple *Threshold rule* where the relationship between seats and votes is linear, but the loser of the election is required to reach a given vote threshold

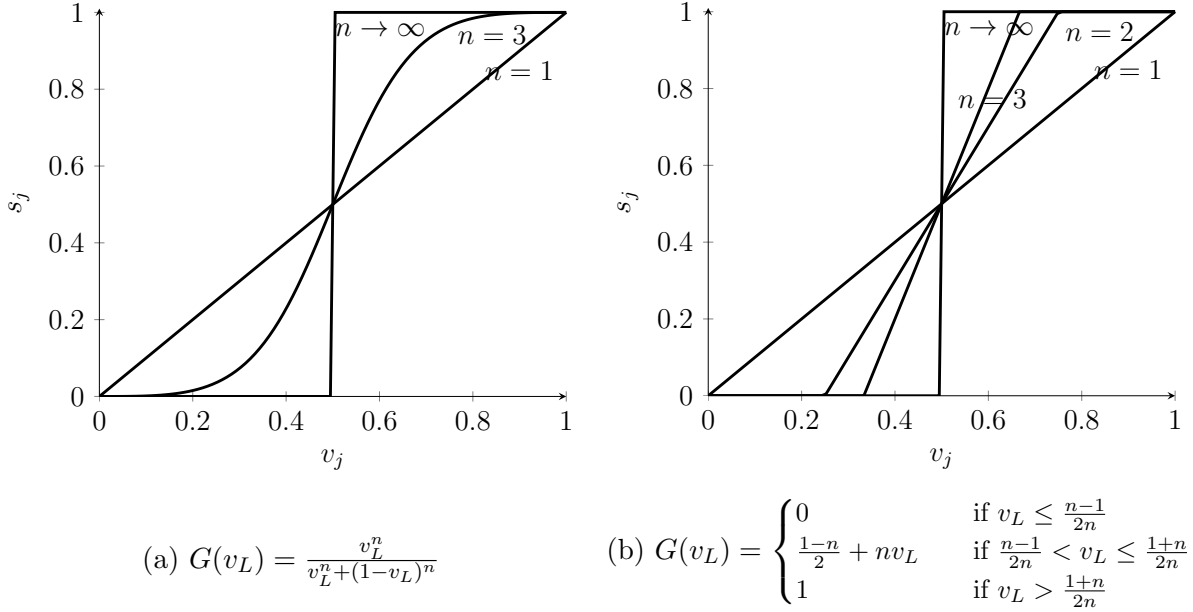


Figure 1: Seat share allocation given parties' vote shares according to two alternative electoral institutions. **Theil's rule** on the left, a **Threshold rule** on the right.

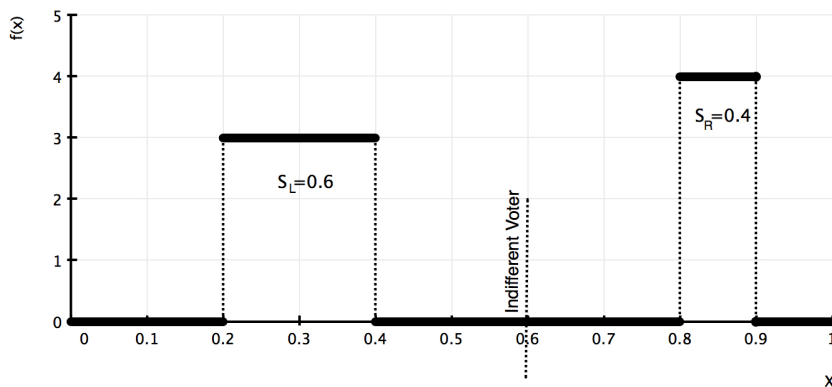
$(n - 1)/2n$ to obtain representation. Formally,

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

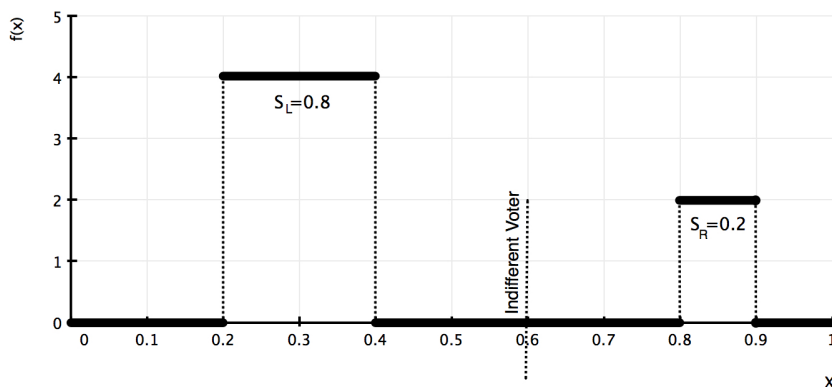
where $n \geq 1$. In Figure 1b, we present the two extreme such institutions in the absence of such threshold leading to a pure PR system ($n = 1$) and that of a winner-take-all election ($n \rightarrow \infty$ and an almost 50% threshold). Intermediate cases with a 1/4 ($n = 2$) and 1/3 ($n = 3$) thresholds are also presented. Clearly, the larger this threshold (large n), the more favouring is the system towards the winner of the election.

Given parties seat shares s_L and s_R , we can now determine the distribution of ideologies of the members of parliament. Formally, the distribution of ideologies will have support on $[\underline{x}_L, \bar{x}_L] \cup [\underline{x}_R, \bar{x}_R]$ (the ideological spectrum chosen by the parties), and ideologies will be uniformly distributed within each party with the density given by the following function:

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



(a) Using *Theil's rule* and $n = 1$, we have that $s_L = 0.6$ and $s_R = 0.4$.



(b) Using *Theil's rule* and $n = 3.42$, we have that $s_L = 0.8$ and $s_R = 0.2$.

Figure 2: The distribution of ideologies for $[\underline{x}_L, \bar{x}_L] = [0.2, 0.4]$ and $[\underline{x}_R, \bar{x}_R] = [0.8, 0.9]$ and hence $v_L = 0.6$ and $v_R = 0.4$.

For an illustration of the above density function and the ideologies represented in the parliament for different levels of disproportionality according to *Theil's rule*, let us refer to Figure 2. Given parties' policy proposals $[\underline{x}_L, \bar{x}_L] = [0.2, 0.4]$ and $[\underline{x}_R, \bar{x}_R] = [0.8, 0.9]$, the indifferent voter is located at 0.6 and hence $v_L = 0.6$ and $v_R = 0.4$. The top panel presents the case of a pure PR system ($n = 1$) and the lower panel presents the case of a system favoring the winning, leftist party ($n = 3.42$). As it is clear, the ideologies included in the list of the leftist party are better represented in the parliament when the system is favoring the winner of the election.

Parties' Payoffs: We assume that parties' payoffs depend on the distribution of ideologies in the constituted parliament. In particular, let party j value each seat representing ideology t by the following expression:

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

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

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

or given the properties of the electoral rule and the distribution of ideologies in parliament according to $f(x)$:

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

where $s_L = G(\frac{\bar{x}_L + \underline{x}_R}{2})$ and $s_R = 1 - G(\frac{\bar{x}_L + \underline{x}_R}{2})$.

2.1 Theoretical predictions

In our game, parties propose lists to elect a parliament to their liking. The crucial question to understand the equilibrium structure is: what are parties' incentives to propose more or less cohesive lists of candidates? By enlarging their lists towards moderate grounds (i.e., high \bar{x}_L and low \underline{x}_R respectively) parties move the indifferent voter in their favor and, hence, obtain a higher vote share since they become more appealing to moderate voters. Clearly, this effect -and the incentives to obtain a high vote share- increase as the electoral rule favors disproportionally the winner of the election. However, including in the list moderate candidates comes at a cost: the ideologies represented in the parliament may become too centrist and, thus, affect negatively parties' payoffs. Therefore, parties also enlarge their lists towards the extremes (i.e., low \underline{x}_L and high \bar{x}_R), despite not affecting their vote shares since extreme voters were anyway voting for them.

Now that the intuition on parties' incentives is clear, let us present in the following proposition the equilibrium characterization and relevant comparative statics.

Proposition 1. *Let $x^* = \frac{x_L + [2 + x_L(6x_L - 7)]G'(0.5)}{1 + (1 - 2x_L)G'(0.5)}$. There exists a unique symmetric equilibrium where:*

1. $[\underline{x}_L^*, \bar{x}_L^*] = [(3x_L - \min\{x^*, 0.5\})/2, \min\{x^*, 0.5\}]$
2. $[\underline{x}_R^*, \bar{x}_R^*] = [1 - \min\{x^*, 0.5\}, 1 - (3x_L - \min\{x^*, 0.5\})/2]$

3. In equilibrium, $\frac{\partial(\bar{x}_j^* - \underline{x}_j^*)}{\partial G'(0.5)} \geq 0$, 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 policy space. Each party $j \in \{L, R\}$ strategically chooses how far from its ideal point its candidates' list should extend depending on the characteristics of the electoral institution captured by $G'(0.5)$ and its ideal point (x_j). Indeed, larger values of $G'(0.5)$ indicate a more disproportional allocation of seats in favor of the larger party. The crucial comparative static shows that the length of the list is increasing as the rule favors disproportionately the winner of the election (i.e., $\partial(\bar{x}_j^* - \underline{x}_j^*)/\partial G'(0.5) \geq 0$).

To visualize the result but also to understand the equilibrium structure further, let us focus on Figure 3 that presents the result for both examples of electoral institutions previously presented (*Theil's* or the *Threshold rule* lead to the same equilibrium since $G'(0.5) = n$ for both rules, where n measures the electoral rule disproportionality). As the figure shows, the length of both parties' lists is increasing in the electoral rule disproportionality. That is, our equilibrium results show that disproportional electoral systems provide incentives to parties to become less cohesive. This is a consequence of the incentives provided by disproportional electoral systems to parties to increase their vote share.

Notice however that, as our results indicate, enlargement does not occur in a symmetric manner around the party's ideal policy. That is, for every unit of enlargement towards the centre so as to search more votes, each party also enlarges towards the extreme by half unit. Enlargement towards the extremes does counterbalance enlargement towards moderate grounds in terms of ideologies represented in the parliament, but, in the equilibrium, the enlargement towards the extremes should be smaller than the enlargement towards the center.

The above arguments are the ones illustrated in Proposition 1 for any electoral institution G described in our model. That is, for every G and x_L , there exists a unique value x^* that determines parties' lists. The *min* operator appearing in the formal result simply restricts the equilibrium values $[\underline{x}_j, \bar{x}_j]$ 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 in their list more extreme candidates.

2.2 Discussion of our model and robustness

Having presented our main result, we can now discuss our main assumptions. Recall that in our setup, parties propose a set of candidates to maximize their policy-related utility

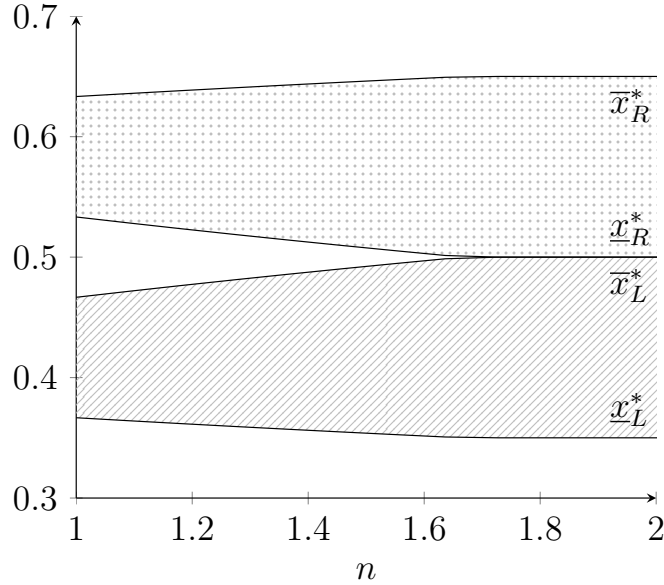


Figure 3: An example of equilibrium intraparty cohesion considering either *Theil's* or the *threshold* rule for different levels of n and ideal policies $(x_L, x_R) = (0.4, 0.6)$. Equilibrium lists coincide for these two rules given that $G'(0.5) = n$ for both.

out of all the elected candidates, while voters vote for the candidate they like the most (as in our empirical setting). Our assumptions can then be seen as the direct extension of the simplest voting model with sincere voters and two policy motivated parties that propose a unique policy (or candidate) to the multi-candidate setting presented. That is, as in the standard model voters sincerely vote for the candidate they like the most and parties care about the policies represented by all the elected candidates. Importantly, several of our assumptions can be relaxed without changing the nature of our main result. While the equilibrium characterization would vary, the main result showing that, as the electoral rule becomes more disproportional parties become less cohesive, would survive. A list of possible modifications follows.

Preferences: Our result is robust to parties having preferences over the mean of the parliament instead of every parliament seat, as we consider here, and to voters caring about the aggregate party's list ideology. Also, one could permit parties to have any symmetric ideal points, and allow the society (and members of parliament) to be distributed in a non-uniform manner. With respect to the distribution of voters, our results are qualitatively identical for any distribution of voters $F(x)$ that is symmetric around $1/2$ and $G(F(x))$ is a log-concave function. Regarding parties' ideologies, currently the restriction is that parties are not too extreme (i.e., $x_L \in [1/3, 1/2]$ and $x_R \in [1/2, 2/3]$). This assumption guarantees that, in the equilibrium, the extreme bounds of the lists will never hit the extremes of the policy space. This might however happen if parties were permitted to be more extreme. Still, this would not affect the nature of our results and

a full characterization of the equilibrium is possible.

Admissible strategies: We assume that parties nominate a uniform distribution of candidates. While, at first sight, this appears as a potential restriction (why should not a party be free to choose more or less polarized distributions of candidates?), it is really equivalent to assuming that parties choose, first, an interval of admissible ideologies, and, then, try to find candidates that fit within (e.g. by interviewing people randomly and nominating the ones that have an admissible ideology). This is arguably quite realistic in many contexts since there is rarely a readily available infinite pool of candidates of all kinds where parties can draw any subset to their liking: Identifying a long list of suitable candidates is a painful process from the parties' perspective (e.g. in our empirical application, parties do not often reach the permitted maximum number of candidates). Hence, assuming essentially that the parties choose only the maximum possible degree of differentiation between candidates seems as the natural way to go. One could, of course, permit parties to choose not only the set of admissible ideologies, as we do, but also their exact distribution. If this additional feature comes at an extra cost (e.g. the party would need to interview more people to generate a more demanding list of candidates), then our main comparative result can be shown to still hold.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, in the Appendix (Section A1.2), we 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 large-scale oversimplifications in the model's assumptions, it is quite evident that the trade-offs that drive the comparative results do not hinge on the exact number of parties involved in electoral competition: If a more inclusive party list increases the electoral performance of a party but it is unappealing policy-wise, then when the electoral rule rewards more an increase in vote-share, a party should expand its list, independently of how many competitors it faces. Our theoretical results need hence be thought as a comparative static of the electoral rule disproportionality on intraparty cohesion for a fixed number of parties.

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

⁶A case that fits this description is when parties –after choosing the interval of admissible ideologies– also choose a pair of parameters (a, b) at a cost $(a - 1)^2 + (b - 1)^2$ to induce a Beta distribution of candidates with parameters (a, b) and a support equal to the selected interval. When $a = 1$ and $b = 1$, the distribution of candidates coincides with the restriction of the distribution of the ideologies of the population in the selected interval, and, thus, it is the cheapest option. When parties want to induce a different distribution of candidates, they need to exert more effort in the selection process. This parametrization of the problem allows for many different kinds of configurations (e.g. unimodal, bimodal, biased to the left/right, symmetric, etc.), and is hence general enough to reassure us of the robustness of the derived results.

to a multiparty system and FPTP systems leading to a two-party system are considered a “law”, and there is indeed supportive evidence. However, the pure PR and FPTP systems are just the two extreme cases of our model as captured by our function $G(v_L, v_R)$. In between these two extremes, we permit continuous 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. Importantly, this seems to be also the setting of our empirical analysis. We focus on PR systems, where although the disproportionality changes via changes in the council size (see Table A1), our balance tests show that there is no statistically significant effect on the number of parties or the effective number of parties (see Table 3).

3 Empirical Evidence

We first describe the institutional details of the empirical setup and then detail our identification strategy. The same identification strategy is then used: a) to present our main results on the effect of the council size on parties’ cohesion, and b) to rule out alternative mechanisms than the electoral rule disproportionality that could potentially explain the effect. In the Appendix (Section A2.2), we also show that indeed our setup provides exogenous variation of the electoral rule disproportionality via changes in the council size. While one may argue that this effect is unsurprising, it is important to demonstrate that this effect is strong enough to show up in our sample for it to be a plausible causal mechanism for any effects on parties’ cohesion.

3.1 Institutional details and link with theory

Although our theoretical model is quite general and does not aim to exactly replicate the voting context of our empirical analysis, it has close parallels on how voting and tallying takes place in Finnish municipal elections that is the focus of our empirical analysis. In each municipality of council size k , parties (or pre-electoral coalitions of parties) propose an open list of up to $1.5*k$ candidates and each voter votes for one candidate. Candidates’ votes are then aggregated at the list level and determine the lists’ vote shares.⁷ Lists’ vote shares are translated into lists’ seats following the D’Hondt allocation method. Seats are in turn allocated to the candidates with the most votes within the list. Thus, similar to our model, parties propose the set of candidates competing in each list and voters vote for a candidate who belongs to one of the lists.⁸ That is, in our empirical setup,

⁷Parties can form pre-election coalitions and propose a joint list of candidates. The allocation of seats then takes place at a coalition list rather than at a party level.

⁸Over 40 countries use an open list PR system.

parties can be seen as proposing a list of candidates that resembles the concept of the interval of ideological heterogeneity $[x_j, \bar{x}_j]$ proposed by each party in our theoretical model. As in our theoretical model, the list composition in reality is also likely to affect council’s ideological composition (and we show that it does), and thus policy outcomes. Finnish local politics do not have very strong party discipline in place - at least, relative to the parliamentary politics in Finland - and even a single individual councillor can substantially affect economic policy (Hyytinen et al. 2018a).

The number of candidates elected in each municipality (i.e., council size) varies between 13 and 85 and is a deterministic step function of the municipalities’ population.⁹ Importantly, while the council size varies seats are allocated following the D’Hondt method in all municipalities. This method, although a member of the “proportional” allocation formulas, is known to be one of the most favorable to large parties (Herron et al. 2018; Gallagher 1991). Even more, and crucially for our setup, the extend of the advantage that the largest party enjoys (in terms of allocated seats) depends on the council size, which varies with the municipality’s population. As the council size grows, the advantage of the large party becomes smaller and hence the electoral rule less disproportional (Herron et al. 2018; Benoit 2000). In the Appendix (Section A2) we present a formal illustration of the effect of council size on the electoral rule disproportionality and further links between our theory and the actual electoral rule in Finland for the interested reader.

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 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 A6 in the Appendix for a detailed description). During the response period, each candidate has access only to her 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 her replies to those of the candidates. The application also provides a list of candidates whose replies are closest to the voter’s. The open-list system makes Finland a fertile ground for the use of the voting

⁹The council sizes for the different population groups are: population less than or equal to 2,000 (council size 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).

aid application as voters must find an individual candidate to vote, so mere party-level information is not enough to guide the choice. Using the application is free of charge for both candidates and voters.¹⁰ We have access to these data only for 2008 and 2012.

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 in total 56.2% of the votes of the municipality. The equivalent figures for 2012 were 47.2% of the candidates and 54.3% of the votes. Generally, the candidates who respond to the vote aid application are politically more successful and experienced, younger and more likely to be women. As we later detail, the response rate is balanced across the cutoffs used in the RDD and hence should not pose any threat to our identification strategy (see Table 3 for these balance tests).

Second, we also 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 data from Statistics Finland's 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 parties pre-electoral coalitions.

Using the electoral data we construct our main disproportionality measures that we detail in Section A2.2. Similarly, using the YLE data, we construct the main outcome variables on parties' cohesion that we detail in section 3.4. All variables are summarized in Table A11 and described in the presentation of balance tests (Section A4).

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.¹¹ The identifying assumption in such RDD is that individuals cannot precisely manipulate the forcing variable (see e.g. Lee

¹⁰Finland 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 favourite candidate and followed the application's recommendation (see Wagner and Ruusuvirta 2012 and references therein).

¹¹Regression discontinuity at population thresholds is a common approach to isolate causal effects. See for example Pettersson-Lidbom (2012); Gagliarducci et al. (2011); Eggers (2015); Bordignon et al. (2016); Ferraz and Finan (2009); Brollo et al. (2013); Fujiwara (2011); Egger and Koethenbueger (2010); Gagliarducci and Nannicini (2013) 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. Similar to us, Sanz (2017) and Lyytikäinen and Tukiainen (2016) use population thresholds to study political consequences of electoral systems.

and Lemieux 2010). This is true in our case, because municipalities do not self-report their population. In this case, identification is based on a local randomization at the threshold.¹²

We are interested mainly in two outcomes. First, we show (in the Appendix Section A2.2) that the council size has the expected effect on the proportionality of the electoral system. 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), as predicted by our theory. 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 and include a flexible but smooth function of population as control variables. The population variables should pick up the impact of all the determinants of within-party cohesion correlated with population, apart from the council size. Hence, we will obtain a reliable estimate of the causal effect of the 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, we report confidence intervals based on the bias-corrected coefficients and the associated robust inference by Calonico et al. (2014) due to the superior coverage properties of the latter method. 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)

¹²However, local randomization is not a requirement but rather one possible interpretation of RDD. The sufficient identification assumption that is that the potential outcomes develop smoothly over the threshold. One difference between these two interpretations of the design is that the latter allows there to be trends in the potential confounders. See Cattaneo et al. (2015) and Sekhon and Titiunik (2017) for further discussion.

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 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 changes also the coefficients because the optimal bandwidths change.

One complication to our analysis is how to deal with multiple thresholds. One 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 for the nonparametric analysis. 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:

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

where Y_{it} is the outcome of interest, v_{it} is the forcing variable measuring the distance from the normalized population cutoffs for each observation i in election t , $1(v_{it} > 0)$ is an indicator function for being above a cutoff and δ is the coefficient of interest. If $f(v_{it})$ 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, δ will be the causal estimate of interest.¹³

However, we also have an interest at the magnitude of the effect at each individual threshold, but, given the sample size, the nonparametric estimation at each cutoff separately is somewhat imprecise. Therefore, we report additional parametric polynomial specifications that rely also on data points further away from the thresholds. However, this parametric approach is mainly used to produce meaningful visualizations of the data.

In all the analysis, we limit our sample to municipalities with a population below 22,500 to focus the analysis around the thresholds where the data is denser. Our sample contains 76% of Finnish municipalities with 31% of population. The results may not apply to larger municipalities, because the council sizes change relatively much less at the larger population cutoffs implying only small changes in proportionality. Moreover, we omit municipalities that underwent a municipal merger prior to elections as these affect council size and many other features of political competition.¹⁴

¹³In 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).

¹⁴However, the cohesion results remain statistically significant if we include the merged units and larger municipalities, but as expected the point estimates are closer to zero (not reported).

Even if our pooling approach is standard in the literature, it is not entirely unproblematic. 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 of the threshold is the same (which would happen in large samples due to local randomization). Therefore, the McCrary (2008) density tests need to be reported separately for each threshold as opposed to the entire pooled sample. We do not observe any jumps neither at any of the individual cut-offs nor at the pooled one (see Figure A4 in the Appendix).

The standard identifying assumptions of our model imply that other possible determinants of intraparty cohesion should develop smoothly with respect to population and be therefore 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 related to the case of analyzing population thresholds, since in many countries, municipalities' responsibilities, grants, politicians' salaries and regulation depend also on the same thresholds. In that case, there are 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 the one proposed by our theory, we rely mainly on the covariate balance tests (see Sections 3.6 and A4 for a detailed discussion and results).

3.4 Main Results: Council Size and Intraparty heterogeneity

As our main outcome variables, we construct two measures of candidate heterogeneity, given the candidates' responses to the voting aid application. The first is constructed using all the available responses (*All questions index*) to avoid selecting on the questions. The second focuses only on a subset of the questions focusing on important economic issues such as taxation and redistribution (*Redistribution index*) (see Section A6 in the Appendix for this selection) and serves for robustness purposes. For both indices, we first compute for each candidate the distance between their own response and the party mean response for each question, and take a square of that. To obtain the index, we aggregate (sum) these squared distances over all the questions included in the index and take a root of the sums of those squares. That is, we use simple Euclidean distances as a measure of ideological heterogeneity.¹⁵ If the distance is zero for a candidate, her

¹⁵There 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

ideology coincides with the party’s mean. The larger this distance is, the more diverse is this candidate compared to the mean. 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. This leaves us with 14999 candidate-election year, 1184 party-election year and 475 municipality-election year observations.¹⁶

We begin the RDD analysis by graphical visualization of the jumps at the cutoff. In Figure 4, we report the results for the two indices of the policy positions using a parametric RDD with a 3rd order polynomial of population. The results are very similar in both cases. Both measures jump down at each of the threshold, but none of the jumps are statistically significant. We report the actual regression results in Tables A4 and A5 in the Appendix for a wider range of different orders of the polynomials.

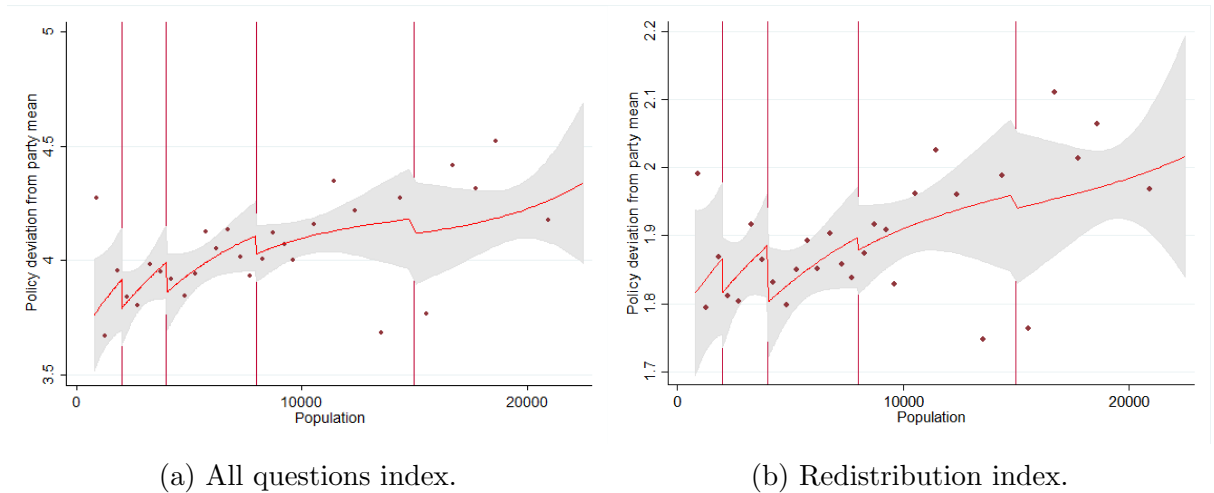


Figure 4: Parametric RDD. 3rd order polynomial.

We present the nonparametric results in Table 1. Overall, the evidence is strongly consistent with our theory: Party cohesion increases (that is, our dependent variable measuring distances decreases) as the council size increases. The estimate is always negative and statistically highly significant in all the cases. We use the individual candidate level data in Table 1. Therefore, clustering 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 local party-year and municipality-year level in Table 2. There, the outcomes are defined as means over relation to its party. Our results are robust also to using either the standardized Euclidean distance or the Mahalanobis distance. These alternative measures account for the differences in the variances across the individual questions.

¹⁶Note that, at the municipal level, we are left with 30 observations less than at the analysis of the disproportionality due to the minimum of five responses we imposed on the YLE data on the candidates’ positions.

the individual candidate distances aggregated to the respective level, and the results are robust.

Table 1: Policy positions and council size, nonparametric RDD (candidate level)

	All questions	Redistribution	All questions	Redistribution
Conventional local linear RD coefficient	-0.427	-0.2824	-0.592	-0.29632
95% Confidence interval with bias-correction and robust inference	[-0.597 ; -0.200]	[-0.431 ; -0.159]	[-1.256 ; -0.128]	[-0.621 ; -0.076]
N within main bandwidth	1466	1865	3115	3112
MSE-optimal bandwidths (main/bias)	247/480	358/529	551/1025	530/1012
Clustered bandwidths and s.e.'s	No	No	Yes	Yes
Notes: Results are generated using rdrobust package in STATA (Calonico et al. 2016).				

Table 2: Policy positions and council size, nonparametric RDD (other levels)

	All questions	Redistribution	All questions	Redistribution
Conventional local linear RD coefficient	-0.302	-0.243	-0.512	-0.267
95% Confidence interval with bias-correction and robust inference	[-0.803 ; 0.071]	[-0.529 ; -0.056]	[-1.174 ; -0.084]	[-0.587 ; -0.059]
Unit of observation	Municipality-year	Municipality-year	Party-year	Party-year
N within main bandwidth	223	180	300	254
MSE-optimal bandwidths (main/bias)	752/1171	608/1037	560/987	493/931
Clustered bandwidths and s.e.'s	Yes	Yes	Yes	Yes
Notes: Results are generated using rdrobust package in STATA (Calonico et al. 2016).				

3.5 Alternative mechanisms

To check for other potential mechanisms through which the council size could affect intraparty cohesion, we conduct several balance tests. In Table 3, we report the most important ones and relegate to the Appendix further tests on municipality and candidates' characteristics (Tables A6-A8). The latter concerns pre-treatment covariates, and, thus, are standard validity tests of RDD, whereas the ones analyzed here are alternative outcomes that the council size could plausibly influence. All the balance tests are reported slightly differently (only coefficient and the associated s.e.) from the main results for the sake of brevity.

First, it is important to stress that, across the cutoffs, the number of candidates and the number of parties (either as the simple count or as the effective number of parties) are balanced, that is, the effect is not statistically significant.¹⁷ The number of parties being balanced is particularly relevant since one could, at first, think that the changes in the (dis)proportionality should also affect the number of parties. Note, however, that here we are 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.

¹⁷As is common in the literature, we compute the effective number of parties (lists) by the inverted Herfindahl index of party lists' vote shares.

Also, in the current context, the parties that are well-known at the national level may not find it worthwhile to merge with other parties and/or other lists across the cutoffs just for small variations in the electoral rule (dis)proportionality. This balanced number of parties is also relevant because, as explained in Section 2.2, our theoretical result is actually applicable to multi-party settings for any exogenous number of parties.

The result that number of candidates is not significantly affected by the rule also highlights that parties and voters act in similar environments across the cut-offs vis-a-vis the list's characteristics. Note that, due to the fact that parties can nominate up to 1.5 times the council size on the list, one could naturally expect jumps across cutoffs. The absence of such jumps can be attributed to the difficulties the parties face in filling the lists in the first place. Notice that lists usually include fewer candidates than the maximum threshold.¹⁸ Moreover, the absence of jumps implies that the increase in the council size does not relax the supply constraints. Therefore, our theoretical focus on the demand side, where parties choose their lists among a set of available candidates seems reasonable in our empirical setting. Furthermore, our main results imply that, while also the supply of candidates seems to be an important constraint, the parties play a crucial gatekeeper role in terms of ideology and very likely with respect to other characteristics, too. One could surely argue that in some other setting, the supply side of candidates could also be affected by changes in the electoral disproportionality. We leave these theoretical considerations for further research.

Finally, we point out that the candidates' response rate to the YLE application is balanced. Given that using the application is voluntary, one could be concerned about a possible selection bias. The balanced rate, however, indicates that a possible selection bias resulting from the response rate is not present in the RDD estimates. Any possible selection bias seems to be the same across the cutoffs, and is thus differenced out from the RDD estimates.

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 council size on these outcomes. To further evaluate the mechanism behind our cohesion results, we show in the Appendix (Table A9) that the disproportionality and the cohesion indices move significantly at the same cutoffs, whereas the alternative mechanisms do not.

3.6 Robustness, validity and discussion

In the Appendix (Section A4), we report and discuss in detail the standard validity and robustness checks. The McCrary (2008) test for manipulation shows no evidence on

¹⁸Only 3.4% of lists are full in our sample. Typically only larger parties in larger municipalities are able to fill the lists completely.

Table 3: Alternative mechanisms

Outcome	Candidates	Numb. of parties	Effective numb. of parties	Candidates	Respondents
Bias-corrected RD coefficient	6.66	0.282	0.010	2.749	0.664
Robust s.e.	7.56	0.402	0.277	2.805	1.093
N within main bandwidth	300	241	245	458	323
Unit of observation	Municipality-year	Municipality-year	Municipality-year	Party-year	Party-year

Notes: Results are generated using `rdrobust` package in STATA (Calonico et al. 2016) using MSE-optimal bandwidth (optimized for each outcome separately) and triangular kernel. Standard errors (and bandwidths) are clustered at the municipality level. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% level.

municipalities manipulating their population count at any individual cutoff nor in the pooled data (Figure A4). This makes perfect sense, because population counts are not self-reported by the municipalities, there are no incentives to manipulate this information, and no other policies or municipality responsibilities change at these cutoffs. We also report that the results are robust across a fair range of bandwidths around the optimal ones (Figure A5).

We report the placebo cutoff analysis in Figure A6 in the Appendix (Section A4). This analysis is especially useful for understanding whether the applied RDD specification is appropriate (Hyytinen et al. 2018b). This analysis further reveals that we should trust the clustered results much more than the non-clustered. This is because there is some within-municipality correlations 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 cutoffs test for the clustered specification works as it is supposed to giving zero results when using the placebo cutoffs.

Finally, we report descriptive statistics of our main variables of interest as well as of the variables used in the covariate balance tests in Table A11. This information will help to further understand the magnitudes of the estimated effects. Regarding our main results on intraparty cohesion, Figure 4 is the most informative in understanding the magnitudes. The effect of crossing the threshold in Table 1 translates into a decrease in the heterogeneity indices by roughly 15 percent relative to the mean value, or half a standard deviation for both cohesion indices.

4 Conclusions

Our work provides new insights on how electoral institutions -and electoral rules in particular- alter parties' incentives when recruiting their political personnel (candidates) and contributes to the literature on candidate selection and nomination (e.g., Besley et al. (2017); Dal Bò et al. (2017); Folke and Rickne (2017)). As Dal Bò and Finan

(2018) note, while parties –and their internal functioning and institutional organization– 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 rules and intraparty ideological cohesion has been largely “black-boxed” in the literature; our study is one of the first that unpacks this link and the mechanism taking place.

Our main result –that more proportional rules generate strong incentives for parties to become more ideologically homogeneous and cohesive– arguably carries non-trivial implications for several other closely related questions. For instance, a logical corollary of Carey and Shugart (1995) is the anticipation that, as within-party competition is intensified in open-list PR the larger the district magnitude is, individual legislators may have incentives to diversify from their colleagues in the face of more intense competition (see e.g., Carroll and Nalepa (2018)). This would then result in parties becoming less ideologically cohesive as the district magnitude increases. Our findings and the theoretical mechanism, however, allow us to reconcile this expectation with recent findings (Cox et al. 2019), because party leadership in open-list PR systems might be using list selection in order to recruit ideologically homogeneous candidates. Thus, in light of the recent increase in partisanship and polarization –at the party-elite level– in the U.S. and Europe, this paper puts forward a potentially relevant reasoning regarding the drivers of voting cohesion and partisanship in legislatures (e.g., Krehbiel and Peskowitz (2015)) and offers new links that connect inter-party polarization with intraparty structure.

Additionally, our arguments offer a rational choice explanation behind the finding by Cox et al. (2016) who show that PR systems are linked with a strong voting coherence by parliamentary parties. In fact, our findings may be seen as an “endogenous” justification: if more proportional systems generate incentives for parties to present more ideologically homogeneous (and thus less diverse) lists, then it is more likely that like-minded legislators will tend to vote in a more coherent manner. Moreover, as the group of legislators becomes more ideologically homogeneous –as our theory predicts– a mechanical reason might kick in: it might be easier for party whips to discipline a more homogeneous group. In other words, our findings can supplement Cox et al. (2016) by unpacking one of the mechanisms justifying such coherence in legislative voting.

Finally, we contribute to the large literature on the electoral rule choice and the trade-off between representation and accountability (e.g. Carey and Hix 2011). Majoritarian rules are considered to favour accountability at the expense of representation, while more proportional rules guarantee better representation (of different voices in parliament) but at the same time make accountability murkier. Our work adds a note of caution to this trade-off. As our model and our empirical results have demonstrated, by generating incentives for more diverse and less homogeneous party-lists, majoritarian systems might

only be offering nominal accountability: while it is true that a single political actor (party) is accountable, the wide variance in possible policy outcomes does not really guarantee that citizens get what they voted for. On the other hand, an analogous caveat is true as far as more proportional systems are concerned. While proportional systems generate incentives for more cohesive and homogeneous parties, thus ensuring that voters get what they voted for, the arguments in favour of them being more representative might have some limits: as the “rubber band” of the list becomes more tight, any gains in representation can only come through the effect of the rule on the number of competing parties (Duverger 1954). Overall, our work demonstrates that the debate on the optimal choice of electoral rules in light of the accountability-representation trade-off may gain additional insights by focusing on within party dynamics. Obviously, our paper does not intend to yield a conclusive verdict on this respect but merely points to an interesting new direction.

Appendix

A1 Theory

A1.1 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 $\int_a^a -(x-t)^2 \frac{1}{a-a} dt = \lim_{b \rightarrow a^+} \int_a^b -(x-t)^2 \frac{1}{b-a} dt = -(x-a)^2$. Hence, for every admissible strategy pair the utility of party L is given by:

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

Notice that, for every \bar{x}_L and \underline{x}_R such that $G\left(\frac{\bar{x}_L + \underline{x}_R}{2}\right) > 0$, we have $\frac{\partial^2 U_L([\underline{x}_L, \bar{x}_L], [\underline{x}_R, \bar{x}_R])}{\partial \underline{x}_L^2} = -\frac{2}{3}G\left(\frac{\bar{x}_L + \underline{x}_R}{2}\right) < 0$ and $\frac{\partial U_L([\underline{x}_L, \bar{x}_L], [\underline{x}_R, \bar{x}_R])}{\partial \underline{x}_L} = \frac{1}{3}G\left(\frac{\bar{x}_L + \underline{x}_R}{2}\right)(3x_L - 2\underline{x}_L - \bar{x}_L)$. 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 $G\left(\frac{\bar{x}_L + \underline{x}_R}{2}\right) = 0$, then, trivially, $\underline{x}_L^*(\bar{x}_L)$ still maximizes the utility of L .¹⁹ 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 $G\left(\frac{\bar{x}'_L + \underline{x}_R}{2}\right) > 0$. Indeed, $U_L\left([x_L, \frac{1}{2}], [\underline{x}_R, \bar{x}_R]\right)$, for example, induces $G\left(\frac{\frac{1}{2} + \underline{x}_R}{2}\right) > 0$ 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 parliament. 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 in the parliament (because $G\left(\frac{\bar{x}'_L + \underline{x}_R}{2}\right) < G\left(\frac{x_L + \underline{x}_R}{2}\right)$) and all the parliament members that it elects are better according to its policy preferences. This means that we

¹⁹The 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 $G\left(\frac{\bar{x}_L + \underline{x}_R}{2}\right) = 0$ does not pose any threat to our equilibrium uniqueness arguments, since in a symmetric equilibrium $G\left(\frac{\bar{x}_L + \underline{x}_R}{2}\right) = \frac{1}{2}$.

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.

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,

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

By log-concavity of G 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

$$\frac{\partial U_L([\underline{x}_L^*(\bar{x}_L), \bar{x}_L], [\underline{x}_R, \bar{x}_R^*(\underline{x}_R)])}{\partial \bar{x}_L} \Big|_{\bar{x}_L = \bar{x}_L^*, \underline{x}_R = 1 - \bar{x}_L^*} = 0 \implies \bar{x}_L^* = \frac{-x_L - 2G'(1/2) + 7x_L G'(1/2) - 6x_L^2 G'(1/2)}{-1 - G'(1/2) + 2x_L G'(1/2)}.$$

Since, $\frac{-x_L - 2G'(1/2) + 7x_L G'(1/2) - 6x_L^2 G'(1/2)}{-1 - G'(1/2) + 2x_L G'(1/2)} < \frac{1}{2}$ if and only if $G'(1/2) \in [1, \frac{1}{3-6x_L})$, 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 $\bar{x}_L^* = \frac{-x_L - 2G'(1/2) + 7x_L G'(1/2) - 6x_L^2 G'(1/2)}{-1 - G'(1/2) + 2x_L G'(1/2)}$ if $G'(1/2) \in [1, \frac{1}{3-6x_L})$ and such that $\bar{x}_L^* = \frac{1}{2}$ if $G'(1/2) > \frac{1}{3-6x_L}$; and $\underline{x}_L^* = \underline{x}_L^*(\bar{x}_L^*)$.

A1.2 Multiparty setting

The incentives to increase the list's heterogeneity as the rule becomes more disproportional presented in our main result do not hinge on assuming a two-party model. To see why, consider, for instance, a set of three parties, $M = \{1, 2, 3\}$, and define by $G_i(v)$ 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:

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

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:

$$\sum_{j \in M} \frac{\partial G_j(v)}{\partial d} \int_{a_j}^{b_j} -\frac{1}{b_j - a_j} (\frac{1}{2} - t)^2 dt - \frac{2}{3} d G_2(v).$$

Considering that the electoral rule is anonymous, we must have $G_j(v) = \frac{1}{3}$ for every $j \in M$, $\sum_{j \in M} \frac{\partial G_j(v)}{\partial d} = 0$ and $2 \frac{\partial G_1(v)}{\partial d} = 2 \frac{\partial G_3(v)}{\partial d} = -\frac{\partial G_2(v)}{\partial d} < 0$. If the rule changes from G to some other anonymous rule, \hat{G} , with $\frac{\partial \hat{G}_2(v)}{\partial d} > \frac{\partial G_2(v)}{\partial d}$ (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 in the paper qualifies to more general setups.

A2 Council size and electoral rule disproportionality

A2.1 Theoretical Arguments

Focusing on the mechanics of the electoral rule of our empirical setting (the D'Hondt method), Figure A1 illustrates for a two party scenario: a) why the advantage of the large party gets smaller as the council size grows (main element for our identification) under this method, and b) how the threshold rule presented in our theoretical model links with our empirical analysis. On the left, we present the seat allocation according to the D'Hondt formula in a council of size 3 and on the right in a council of size 7. In both panels, the step function represents the actual seat share obtained by each party in the council according to the D'Hondt formula for different hypothetical vote shares. To see how our model links to the actual allocation, focus on the solid line that passes through the midpoint of each step. This is exactly the function we used in the example of the

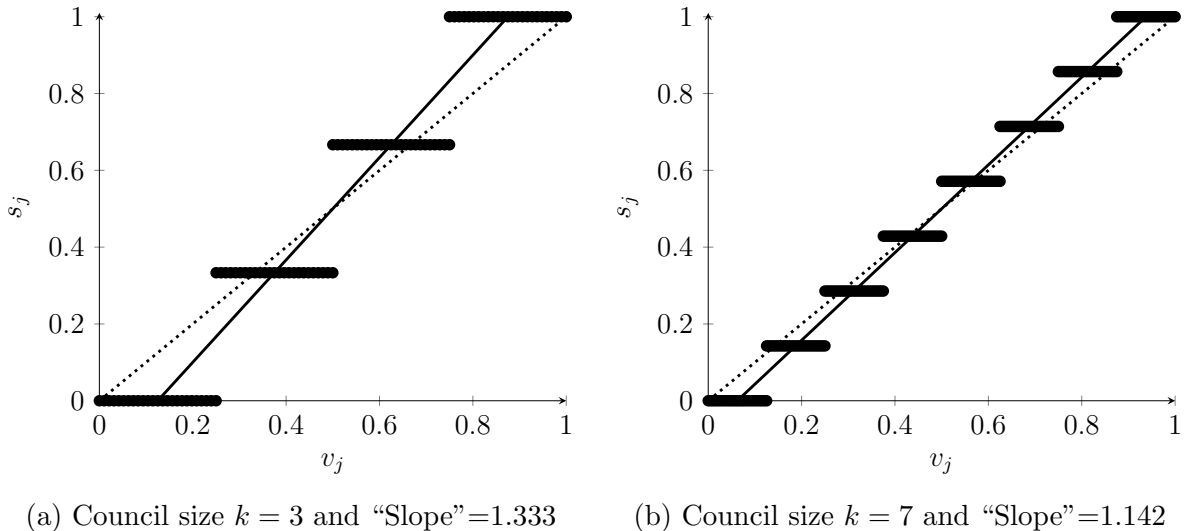


Figure A1: D’Hondt allocation method (steps) in a council of size k and its continuous approximation using the *threshold method* (solid line) where the first seat is obtained when $v_L > 1/2(k + 1)$ and $G(v_L) = -\frac{1}{2k} + (1 + \frac{1}{k})v_L$ (for the increasing part). Dotted line is the 45 degree line (i.e., pure PR: $n = 1$ or $k \rightarrow \infty$).

threshold rule and, as Figure A1, shows can be seen as a continuous approximation of the actual D’Hondt method. The slope of this function, $(1 + 1/k)$, now depends on the council size, k .

What one can also observe is that, in both panels, the party that wins the election (i.e., obtains at least 50% of the vote share) tends to be favoured by the electoral system. As both panels show, for $v_j > 50\%$, the solid line lies above the dashed line meaning that the winner tends to be favored regardless of the council size. Note however, that the the solid line becomes flatter as the council size increases. Hence, the winner of the election is favored more in *smaller-size municipalities*, implying that using the same allocation method in smaller councils tends to be *more disproportional* than in larger ones. That is, council size k is the mirror image of our *threshold rule* parameter n : as k grows, disproportionality n decreases and the allocation of seats is less favorable for the large party.

The above arguments illustrate how our choice of modeling the electoral rule disproportionality through the threshold rule serves as a continuous approximation of the D’Hondt method in a two-party scenario. In reality, Finland has a multi-party system, and, as well known, the number of parties is an essential feature of different electoral systems (Duverger 1954).²⁰ Importantly, the arguments made regarding the D’Hondt

²⁰Currently, 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

method favouring larger parties are also valid in a multiparty context (Herron et al. 2018; Gallagher 1991).²¹

A2.2 Supporting Evidence

In this section, we analyze how the council size affects proportionality in our data. Our data support the theoretical arguments and large parties are favored disproportionately as the council size becomes smaller. The unit of observation is a municipality in a given year. If parties have formed a pre-election coalition, and thus, run as a single joint list in the elections, we define that as a single party when calculating the proportionality measures. This is to best reflect the actual election mathematics and is thus the only sensible choice for the proportionality analysis. However, when analyzing party cohesion, both coalition level and party label level analysis would make sense. For consistency, we use the same unit (coalition) in both analyses.²²

The early debate on the best way to measure the electoral system disproportionality is still open (e.g., Lijphart 1995). In our analysis, we use one existing measure for proportionality and we introduce another. To calculate these, we use 2008 and 2012 election data on all the municipalities with population below 22,500. In total, we have 505 observations at the municipality year level for which we compute the disproportionality measures as our main dependent variable.

One of the most common ways of measuring distortions created by the electoral system is the Gallagher index (Gallagher 1991). The *Gallagher index* in municipality i in year t is defined as:

$$G_{it} = (1/2 \times \sum_{j=1}^p (s_j - v_j)^2)^{1/2}$$

where $j = 1, \dots, p$ denotes the p different parties running in municipality i in year t . The difference $s_j - v_j$ represents the distortions created by the electoral system when a party j that obtains vote share v_j is allocated a seat share s_j . In a pure PR system where no distortions are present (in the examples of the Theil or threshold rule, $n = 1$) this difference takes value zero for each party and so is the case for the index. As the

shares between municipalities.

²¹The interested reader can refer to Herron et al. (2018) for illustration of these arguments for multiparty elections. What is crucial to note is that, while indeed the divisibility of seats affects the proportionality (in the extreme case where the council size is one we are in a first-past-the-post system), different allocation formulas tend to be “less” or “more” proportional. Indeed, the D’Hondt method is known to favor the large parties as opposed to, say, the largest remainder method.

²²These coalitions are formed solely for the purpose of election mathematics favoring larger parties, and they are not taken into account in the actual policy making in the council. Therefore, the RDD analysis on the party cohesion could as well be conducted at the party label level. While we report the analysis only at the coalition level, the results are similar at party label level. Roughly 15% of the lists are such coalitions.

distortions start getting larger the value of the index is also increasing.

Despite the attractiveness of the Gallagher index being its intuitive meaning and ease of calculation, Taagepera and Grofman (2003) argued that it fails to satisfy some relevant axiomatic properties that other indices achieve (e.g., Dalton’s principles of transfers, scale invariance, orthogonality). We therefore use the *Modified Gallagher index*.²³ The *Modified Gallagher index* in municipality i in year t is defined as:

$$MG_{it} = (1/2 \times \sum_{j=1}^p (\frac{s_j}{\sum_{j=1}^p (s_j^2)^{1/2}} - \frac{v_j}{\sum_{j=1}^p (v_j^2)^{1/2}})^2)^{1/2}$$

Again, this index takes value zero in the case of pure PR and higher values in the presence of distortions.

Notice, however, that while the *Modified Gallagher index* (as all others in the literature) represents the level of distortions in the vote to seat share translation, it remains silent on the direction of these distortions. That is, it does not permit us to understand whether such distortions favor the small or large parties, an element crucial to our theory. To be able to capture the direction of such distortion, we propose the use of the *Slope index* constructed as follows: For each municipality-year combination observation, we regress the difference $s_j - v_j$ on v_j . Then, we define the slope of the line obtained from such regression as the *Slope index*. Effectively, it relates the vote share of the parties and their advantage or disadvantage in translating the votes to the seats.

Figure A2 illustrates how the slope of such line captures not only the size of such distortions but also the direction. On the left we depict one municipality-year observation for which the differences are very small, and the slope of such regression is quite flat (0.011). This flat slope indicates the absence of large distortions (in a pure PR system the slope would be zero). On the right, we depict another observation for which the slope is positive and relatively large (0.27), pointing at the electoral system favoring the larger parties. The slope of the line used as our *Slope index* indicates whether systems favor large parties (positive slope), small parties (negative slope) or do not impose any distortions (flat line). Hence, again, the index for a pure PR system is zero while a positive value imply a disproportional system in favor of the large parties.

Next, we analyse whether these measures jump at the population cutoffs that determine the council size. We begin this RDD analysis by a graphical visualization of the jumps at the cutoff. In Figure A3, we report the results for the two different indices of disproportionality using a parametric RDD with a 3^{rd} order polynomial of population (see Section A3 for details). While this specification is quite inflexible, it is informative

²³Koppel and Diskin (2009) formalized the analysis by Taagepera and Grofman (2003) and actually showed that the modified version of the Gallagher index satisfies all relevant properties.

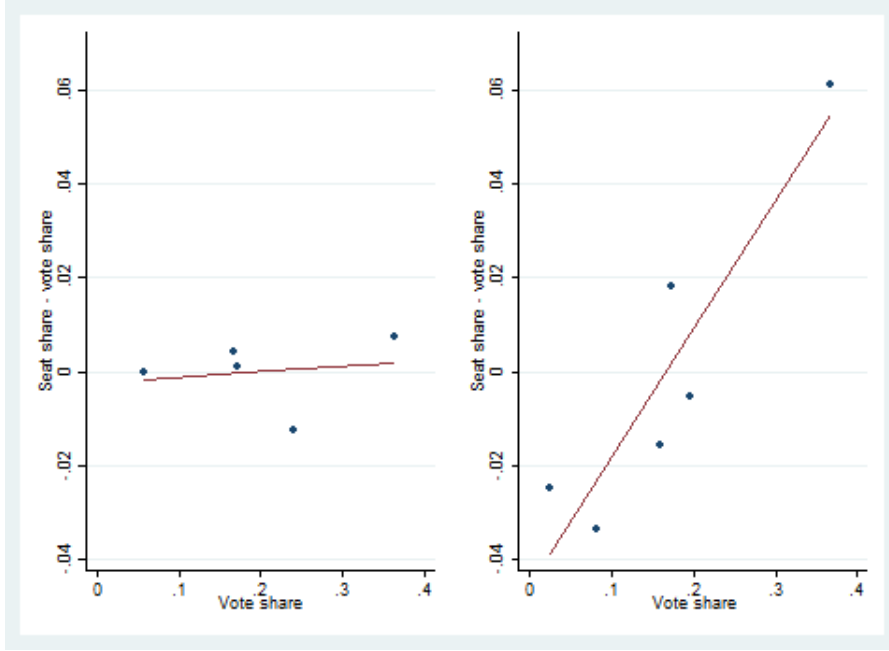


Figure A2: The *Slope index* as the slope of the regression of $s_j - v_j$ on v_j . The slope index takes value 0.0113 on the left (Ilmajoki municipality in year 2012 with 5 competing lists and council size 35) and 0.2712 on the right (Utsjoki municipality in year 2008 with 6 competing lists and council size 21).

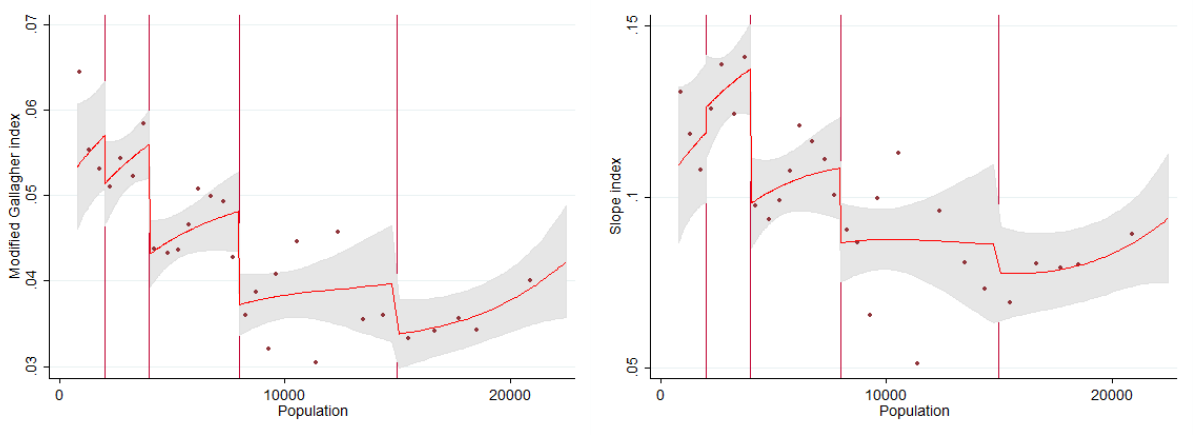
of the slopes over the population and the jumps at each individual threshold. The results are very similar for both indices. Both jump down at each threshold with the largest jumps occurring at the second and third thresholds. We report the actual regression results in Tables A2 and A3 for a wider range of different orders of the polynomials.

In Table A1, we report the nonparametric RDD results on the effect of the council size on proportionality. As for the main results, we report the conventional local linear MSE-optimal coefficient, and, for statistical inference, we report confidence intervals based on the bias-corrected coefficient and the associated robust inference by Calonico et al. (2014) due to its superior coverage properties. We also report both the non-clustered results and those clustered at the municipality level. In line with our theory, the negative coefficients imply that the elections become more proportional as the council size increases. The results are statistically significant at 5% or 10% level, depending on the index.

A3 Parametric RDD results

In this section, we report the cohesion results using the parametric RDD. We estimate by OLS the following equation:

$$y_{it} = \beta_1 + \beta_2 \text{Group}2_{it} + \beta_3 \text{Group}3_{it} + \beta_4 \text{Group}4_{it} + \beta_5 \text{Group}5_{it} + f(\text{Pop}_{it}) + e_{it}.$$



(a) Modified Gallagher index.

(b) Slope index.

Figure A3: Parametric RDD. 3rd order polynomial.

Table A1: Proportionality and council size, nonparametric RDD

	Slope index	
Conventional local linear RD coefficient	-0.027	-0.027
95% Confidence interval with bias-correction and robust inference	[-0.068 ; 0.005]	[-0.065 ; 0.002]
N within main bandwidth	274	271
MSE-optimal bandwidths (main/bias)	838/1460	820/1465
Clustered bandwidths and s.e.'s	No	Yes
	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	238	239
MSE-optimal bandwidths (main/bias)	713/1098	722/1087
Clustered bandwidths and s.e.'s	No	Yes

Notes: Results are generated using rdrobust package in STATA (Calonico et al. 2016).

The dependent variable is the respective index in municipality i in election year t . Function f is a polynomial of population. We use 1st – 7th order polynomials.

The explanatory variables of interest are overlapping dummies $Group2$, $Group3$, $Group4$, $Group5$, indicating all the municipalities above a given threshold. For example, $Group2$ includes all the municipalities with a population of more than 2000. Our estimating sample contains data from the first five groups, because we limit the analysis to the municipalities with a population of less than 22500 to keep the data dense. The respective group coefficients $\beta_2, \beta_3, \beta_4, \beta_5$ give direct estimates of the effect on the index of increasing council size by one step. Thus, the group dummies can be interpreted as individual treatment variables, with the previous group as the control group. Therefore, this specification allows for a different effect at each threshold.

The main drawback of this kind of a parametric RD model is that it uses data far from

the cut-offs to estimate the value of the polynomial at the cut-off. The average effect is calculated as a weighted (by number of observations around each cut-off) average of the individual jumps.

In the case of the proportionality indices, we find that the negative average effect is fairly consistent across specifications and significant at 5% level or 1% level. The overall effect seems to be driven by the individual jumps at the second and third threshold.

Table A2: Proportionality and council size, parametric RDD for *Slope index*

Dep var: Slope index							
Threshold	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th
pop>2k	0.015 [0.0112]	0.014 [0.0122]	0.008 [0.0143]	0.012 [0.0171]	0.019 [0.0187]	0.020 [0.0193]	0.020 [0.0192]
pop>4k	-0.0312*** [0.0071]	-0.0329*** [0.0098]	-0.0379*** [0.0119]	-0.0379*** [0.0119]	-0.0427*** [0.0117]	-0.0441*** [0.0122]	-0.0469*** [0.0129]
pop>8k	-0.0245*** [0.0091]	-0.0259** [0.0117]	-0.0225** [0.0112]	-0.0260* [0.0138]	-0.0263* [0.0140]	-0.024 [0.0159]	-0.023 [0.0163]
pop>15k	-0.017 [0.0150]	-0.015 [0.0132]	-0.007 [0.0162]	0.001 [0.0182]	-0.007 [0.0161]	-0.010 [0.0202]	0.002 [0.0188]
Avg. effect	-0.015** [0.0058]	-0.016** [0.0075]	-0.017** [0.0077]	-0.016** [0.0080]	-0.017** [0.0078]	-0.017** [0.0077]	-0.016** [0.0078]
<i>N</i>	505	505	505	505	505	505	505
adj. R-sq	0.092	0.09	0.09	0.088	0.091	0.089	0.088

Notes: Average effect is calculated as the weighted average of the individual jumps where the weights are based on the number of observations around each cutoff. Standard errors are clustered at the municipality level. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% level.

Table A3: Proportionality and council size, parametric RDD for modified Gallagher index

Dep var: Modified Gallagher index							
Threshold	Order of polynomial of pop						
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th
pop > 2k	-0.003 [0.0036]	-0.004 [0.0038]	-0.006 [0.0044]	-0.005 [0.0051]	-0.003 [0.0055]	-0.002 [0.0056]	-0.002 [0.0056]
pop > 4k	-0.0101*** [0.0022]	-0.0107*** [0.0029]	-0.0123*** [0.0035]	-0.0123*** [0.0035]	-0.0134*** [0.0035]	-0.0144*** [0.0036]	-0.0161*** [0.0037]
pop > 8k	-0.0118*** [0.0029]	-0.0122*** [0.0037]	-0.0111*** [0.0035]	-0.0120*** [0.0043]	-0.0120*** [0.0043]	-0.0108** [0.0049]	-0.0101** [0.0050]
pop > 15k	-0.0089** [0.0045]	-0.0082** [0.0039]	-0.006 [0.0047]	-0.004 [0.0052]	-0.006 [0.0048]	-0.008 [0.0060]	0.000 [0.0056]
Avg. effect	-0.0083*** [0.0018]	-0.0086*** [0.0023]	-0.0088*** [0.0023]	-0.0086*** [0.0024]	-0.0088*** [0.0023]	-0.0090*** [0.0023]	-0.0083*** [0.0023]
<i>N</i>	505	505	505	505	505	505	505
adj. R-sq	0.15	0.149	0.148	0.147	0.148	0.147	0.15

Notes: Average effect is calculated as the weighted average of the individual jumps where the weights are based on the number of observations around each cutoff. Standard errors are clustered at the municipality level. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% level.

Next, we report the results from a parametric RDD for the two policy position indices. Note that, here, the unit of observation is an individual candidate in a given election year.

The results are very similar for both indices. We find that the average effect is consistently negative across the specifications. Moreover, all the individual jumps are negative in all 28 cases reported for both indices. While the results are not statistically significant, the overall pattern is suggestive of a negative jump consistent with our theory.

Table A4: Parametric RDD results, all questions index

Dep. var.: Individual candidate deviation from own party mean, index based on all the questions							
Threshold	Order of polynomial of pop						
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th
<i>pop</i> > 2 <i>k</i>	-0.213*	-0.215*	-0.306*	-0.363*	-0.352*	-0.307	-0.340
	[0.118]	[0.128]	[0.156]	[0.200]	[0.209]	[0.219]	[0.211]
<i>pop</i> > 4 <i>k</i>	-0.067	-0.070	-0.161	-0.181	-0.181	-0.214	-0.168
	[0.086]	[0.115]	[0.150]	[0.150]	[0.150]	[0.169]	[0.185]
<i>pop</i> > 8 <i>k</i>	-0.096	-0.099	-0.103	-0.077	-0.080	-0.047	-0.082
	[0.097]	[0.110]	[0.109]	[0.133]	[0.133]	[0.162]	[0.161]
<i>pop</i> > 15 <i>k</i>	-0.178	-0.177	-0.097	-0.140	-0.140	-0.210	-0.242
	[0.152]	[0.159]	[0.169]	[0.218]	[0.216]	[0.277]	[0.312]
Avg. effect	-0.110	-0.112	-0.117	-0.131	-0.131	-0.144	-0.156
	[0.076]	[0.076]	[0.076]	[0.083]	[0.083]	[0.090]	[0.100]
<i>N</i>	13930	13930	13930	13930	13930	13930	13930
adj. R-sq	0.063	0.063	0.064	0.064	0.064	0.064	0.064

Notes: Average effect is calculated as the weighted average of the individual jumps where the weights are based on the number of observations around each cutoff. Standard errors are clustered at the municipality level. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% level.

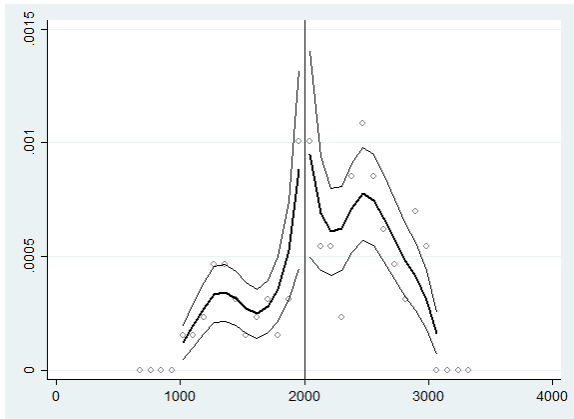
Table A5: Parametric RDD results, redistribution index

Dep. var.: Individual candidate deviation from own party mean, redistribution index							
Threshold	Order of polynomial of pop						
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th
<i>pop</i> > 2 <i>k</i>	-0.102*	-0.104	-0.128	-0.161*	-0.138	-0.136	-0.139
	[0.061]	[0.066]	[0.078]	[0.096]	[0.103]	[0.106]	[0.105]
<i>pop</i> > 4 <i>k</i>	-0.073*	-0.075	-0.100	-0.111	-0.111	-0.113	-0.108
	[0.042]	[0.055]	[0.070]	[0.070]	[0.070]	[0.079]	[0.087]
<i>pop</i> > 8 <i>k</i>	-0.037	-0.039	-0.040	-0.025	-0.032	-0.030	-0.034
	[0.050]	[0.056]	[0.056]	[0.067]	[0.067]	[0.079]	[0.080]
<i>pop</i> > 15 <i>k</i>	-0.058	-0.058	-0.036	-0.061	-0.060	-0.064	-0.067
	[0.076]	[0.079]	[0.088]	[0.109]	[0.108]	[0.140]	[0.155]
Avg. effect	-0.051	-0.052	-0.054	-0.061	-0.061	-0.062	-0.063
	[0.038]	[0.038]	[0.038]	[0.042]	[0.041]	[0.045]	[0.049]
<i>N</i>	14496	14496	14496	14496	14496	14496	14496
adj. R-sq	0.035	0.035	0.035	0.035	0.035	0.035	0.035

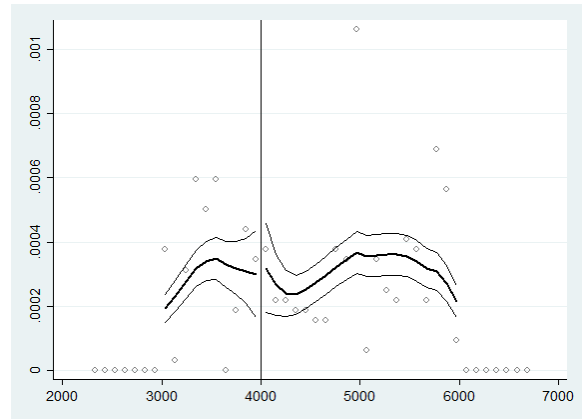
Notes: Average effect is calculated as the weighted average of the individual jumps where the weights are based on the number of observations around each cutoff. Standard errors are clustered at the municipality level. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% level.

A4 Robustness and validity checks

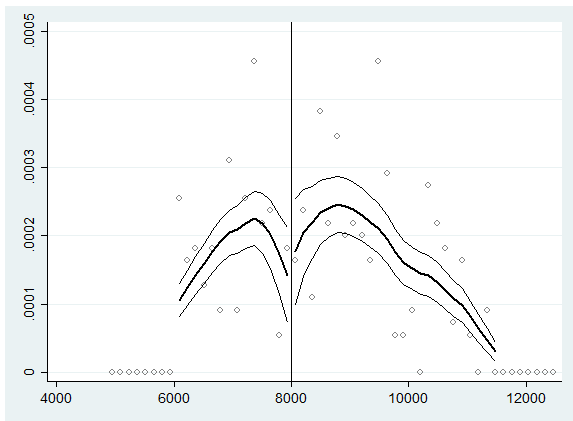
First, in Figure A4, we present the **McCrary (2008) test for manipulation** separately for each threshold as well as for the pooled data. There is no indication of municipalities sorting across the cutoffs.



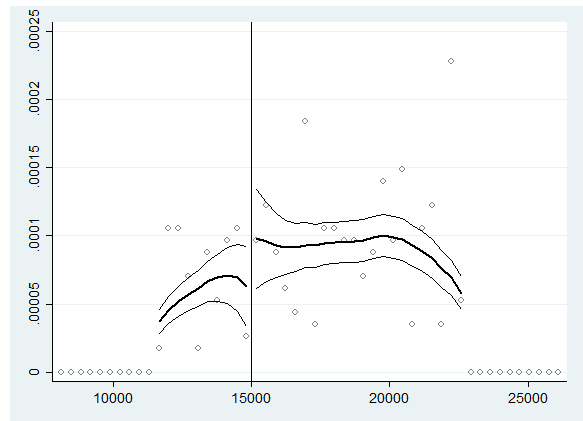
(a) 2000 population cutoff



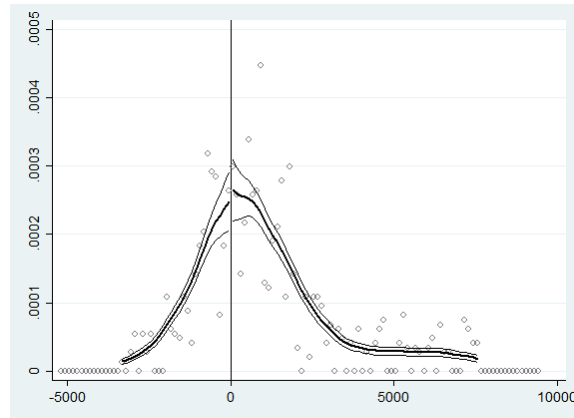
(b) 4000 population cutoff



(c) 8000 population cutoff



(d) 15000 population cutoff



(e) Pooled population cutoff

Figure A4: McCrary density tests.

Second, in Figure A5, we report our analysis of the robustness of the nonparametric main results to the **bandwidth choice**. We report both the clustered and non-clustered results for both the policy position indices. The results are robust across a fair range of bandwidths.

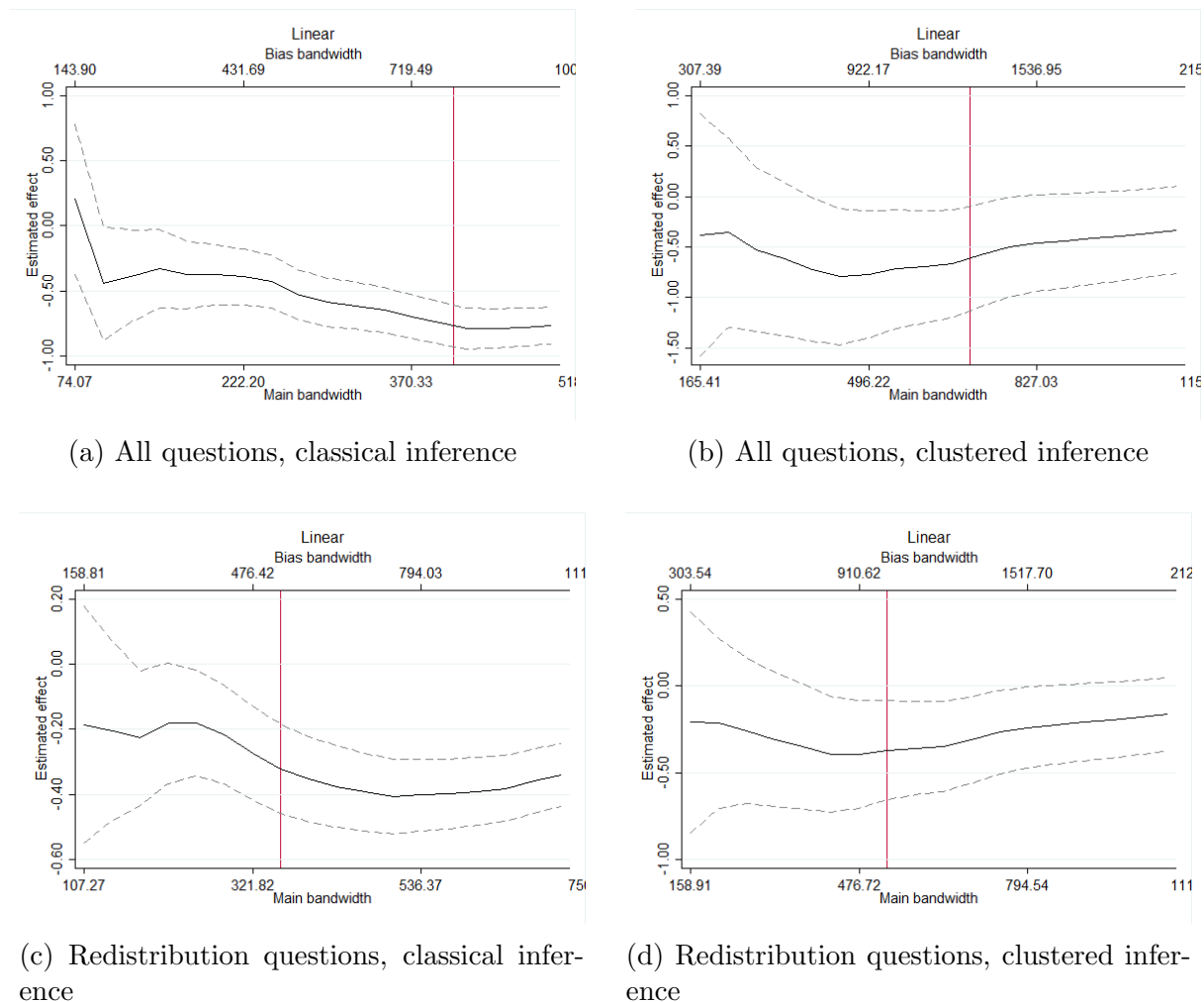


Figure A5: Robustness of the results for alternative bandwidths.

Third, in Figure A6, we conduct the **placebo cutoff analysis**. Here, we artificially move the cutoffs away from their real location. The x -axis shows 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 percent confidence interval. If the design is valid and the specification appropriate, we should observe that the placebo coefficients are not statistically different from zero. This analysis is especially useful for understanding whether the applied RDD specification is appropriate (Hyytinen et al. (2018b)).

We observe that the non-clustered results show a lot of significant positive and negative coefficients. This analysis reveals that we should not trust the non-clustered results. This is because there is some within-municipality correlations 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, and 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.

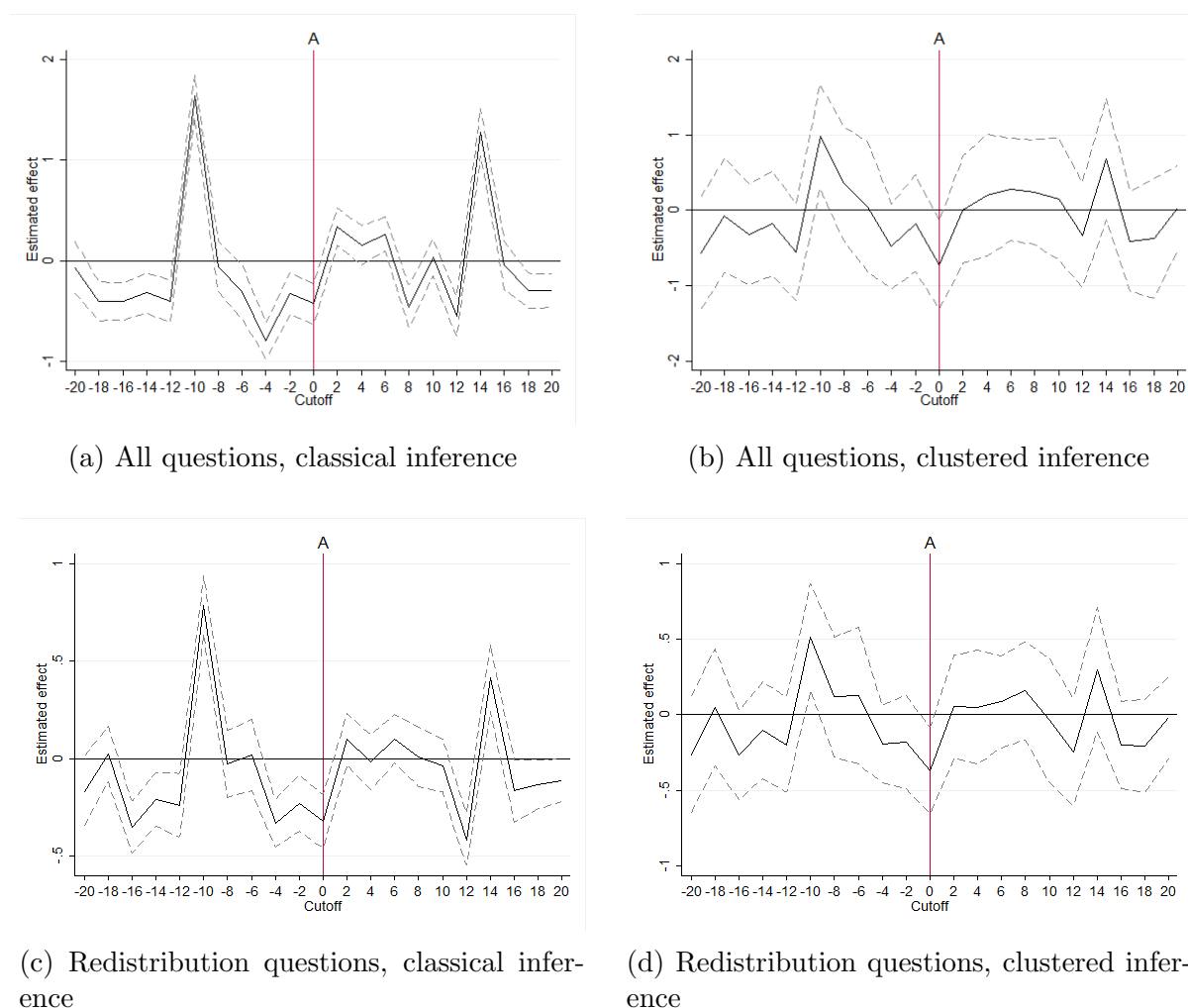


Figure A6: Robustness of the results for placebo cutoffs.

Fourth, in Tables A6-A8, we present further **balance tests** on additional variables compared to the ones presented in the main text. We report the parametric results, and both the clustered and non-clustered nonparametric results. To maintain comparison with the parametric results, we only report the bias-corrected point estimates and robust standard errors in the nonparametric results. We focus here attention on the variables that are realized before the elections and, thus, can be seen as pre-treatment covariate balance tests. Lyytikäinen and Tukiainen (2016) analyze with larger data (more election years and more cutoffs) the effect of the council size on turnout and pivotal probabilities. We do not study these outcomes as they are post-election outcomes that can be influenced by cohesion but not the other way around.

Table A6: Covariate balance tests, parametric RDD

Panel A: Economic and population characteristics of the municipalities							
Outcome	Personnel	Taxes	Over 65yo	Grants	Expen.	Unemp.	Council Size
Avg. effect	46.7	0.0304	-0.00633	-0.0826	-141	-0.197	6.17***
s.e.	135	0.0795	0.0085	0.135	196	0.718	0.0282
<i>N</i>	475	475	475	475	475	475	475
Panel B: Candidate level characteristics							
Outcome	Unemployed	University	Male	Old	Incumbent		
Avg. effect	-0.01	0.0159	-0.00987	-0.00592	0.0283**		
s.e.	0.00602	0.0151	0.0119	0.0127	0.0119		
<i>N</i>	14999	14999	14999	14999	14999		

Notes: Results are from a parametric RDD using 3rd order polynomial specification. The average effect is calculated as the weighted average of the individual jumps where the weights are based on the number of observations around each cutoff. Standard errors are clustered at the municipality level. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% level.

Table A7: Covariate balance tests, CTT

Panel A: Economic and population characteristics of the municipalities							
Outcome	Personnel	Taxes	Over 65yo	Grants	Expen.	Unemp.	Council Size
RD coefficient	-345	-0.069	0.006	-0.260	-428	-1.249	6.28**
s.e.	269	0.093	0.013	0.218	360	1.229	1.60
<i>N</i> within main bandwidth	210	207	314	221	298	243	230
Panel B: Candidate level characteristics							
Outcome	Unemployed	University	Male	Old	Incumbent		
Avg. effect	0.004	-0.005	-0.012	-0.038	-0.060**		
s.e.	0.0004	0.027	0.034	0.023	0.030		
<i>N</i> within main bandwidth	4538	4418	5874	6616	6189		

Notes: Results are generated using rdrobust package in STATA (Calonico et al. 2016) using MSE-optimal bandwidth (optimized for each outcome separately) and triangular kernel. Standard errors (and bandwidths) are classical. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% 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 rate (*Taxes*),

Table A8: Covariate balance tests, CTT, clustered

Panel A: Economic and population characteristics of the municipalities							
Outcome	Personnel	Taxes	Over 65yo	Grants	Expen.	Unemp.	Council Size
RD coefficient	-341	-0.070	0.006	-0.260	-446	-1.26	6.23***
s.e.	308	0.101	0.016	0.274	409	1.43	2.02
<i>N</i> within main bandwidth	220	208	314	323	295	246	241
Panel B: Candidate level characteristics							
Outcome	Unemployed	University	Male	Old	Incumbent		
Avg. effect	0.003	-0.004	-0.010	-0.039	-0.060**		
s.e.	0.012	0.030	0.031	0.031	0.036		
<i>N</i> within main bandwidth	6541	4071	6986	6986	6823		

Notes: Results are generated using `rdrobust` package in STATA (Calonico et al. 2016) using MSE-optimal bandwidth (optimized for each outcome separately) and triangular kernel. Standard errors (and bandwidths) are clustered at the municipality level. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% level.

share of citizens over the age of 65 (*Over 65yo*), central government transfers in 1000€ per capita (*Grants*), expenditures in € per capita (*Expen.*) and the unemployment rate (*Unemp.*). All these variables are measured in the election year and, thus, cannot yet be affected by the council that is elected in that 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. Based on the parametric RDD, more incumbents run as council size increases. This is potentially a concern as the incumbency status is also positively correlated with having preferences closer to the party mean (not reported). However this seems to be simply about due to moderate candidates getting more votes (Meriläinen and Tukiainen 2016) as getting elected does not seem to change policy positions in Finland (Savolainen 2018). Moreover, having more incumbents should be simply mechanical: There are more incumbents in larger councils by construct. Thus, it does not seem that candidates of different competence are selected to the lists across the cutoffs (see also that the other candidate characteristics balance). Therefore, incumbency unbalance is very unlikely to be driving our cohesion results. Moreover, the result is not robust to nonparametric RDD, where we, surprisingly, see a negative and significant coefficient, which is difficult to rationalize and may be just a multiple testing fluke.

However, to argue further that incumbency status does not confound the results, we study in Table A9 whether incumbency changes most at those cutoffs where proportionality changes 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 proportionality changes. However, we show in the same table that their coefficients are not statistically significant, and cohesion is the only variable that shows a significant effect at the same cutoff as proportionality. Thus, the overall evidence gives the strongest support to proportionality being behind the cohesion response.

Table A9: Mechanisms

	pop.>2k	pop.>4k	pop.>8k	pop.>15k
Modified Gallagher	-0.01	-0.02***	-0.02***	-0.02
Slope index	0.02	-0.10***	-0.06**	-0.08
All questions index	-0.32	-0.32	-1.10**	0.92
Redistribution index	-0.12	-0.37	-0.52*	0.64
Incumbency	-0.13	-0.20	0.02	-0.44***
Candidates (munic.)	-1.58	1.76	3.76	-6.09
Numb parties	-0.76*	-0.30	0.44	-1.77
Eff numb par	-0.72*	0.15	0.01	-1.14**
Candidates (party)	1.52	11.5	9.14	43.85*
Respondents (party)	-3.35**	-2.01	-0.51	-1.98

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. Standard errors (and bandwidths) are clustered at the municipality level. *** denotes statistical significance at 1% level, ** denotes statistical significance at 5% level and * denotes statistical significance at 10% level.

Fifth, in Table A10, 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 the municipality grows larger. That is, our main effect of disproportionality affecting parties' selection incentives when recruiting candidates for their lists seems to carry over to the elected body.

Table A10: Elected candidates only: policy positions and council size, nonparametric RDD (candidate level)

	All questions	Redistribution	All questions	Redistribution
Conventional local linear RD coefficient	-0.393	-0.277	-0.454	-0.299
95% Confidence interval with bias-correction and robust inference	[-0.625 ; -0.153]	[-0.479 ; -0.124]	[-1.021 ; -0.059]	[-0.643 ; -0.063]
N within main bandwidth	860	961	1522	1257
MSE-optimal bandwidths (main/bias)	361/547	375/644	554/1023	475/943
Clustered bandwidths and s.e.'s	No	No	Yes	Yes

Notes: Results are generated using rdrobust package in STATA (Calonico et al. 2016).

A5 Descriptive statistics

Table A11: Descriptive statistics

Panel A: Economic and population characteristics of the municipalities							
	Personnel	Taxes	Over 65yo	Grants	Expen	Unemp	Population
mean	2271	2.8	0.22	2.3	6300	11.9	6846
s.d.	774	0.5	0.05	0.8	1094	3.7	5143
Panel B: Political characteristics of the municipalities							
	Council size	Candidates	Numb parties	Eff numb par	Turnout	Redistribution	All questions
mean	27.6	81.0	5.91	3.49	0.65	1.85	3.94
s.d.	7.7	38.5	1.44	0.92	0.05	0.38	0.80
Panel C: Party level characteristics							
	Vote share	Seat share	Respondents	Candidates	Redistribution	All questions	
mean	0.28	0.29	12.7	14.9	1.87	3.99	
s.d.	0.16	0.18	6.5	12.0	0.42	0.83	
Panel D: Candidate level characteristics							
	Unemployed	University	Male	Old	Incumbent	Redistribution	All questions
mean	0.04	0.19	0.58	0.15	0.28	1.90	4.07
s.d.	0.19	0.40	0.49	0.36	0.45	0.72	1.11

A6 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 four-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.

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.
- Lets 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.
- The following questions have a four-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.
 - The five-year long dismissal period for the municipal employees in conjunction with a municipality mergers is too long.
 - Municipal employees should not be nominated as municipal board members.

References

- ACEMOGLU, D., A. OZDAGLAR, AND E. YILDIZ (2011): “Diffusion of innovations in social networks,” *Proceedings of the 50th IEEE Conference on Decision and Control*.
- AGHION, P. AND J. TIROLE (1997): “Formal and real authority in organizations,” *Journal of Political Economy*, 105, 1–29.
- BEAL, D. J., R. R. COHEN, M. J. BURKE, AND C. L. MCLENDON (2003): “Cohesion and performance in groups: A meta-analytic clarification of construct relations,” *Journal of Applied Psychology*, 88, 989–1004.
- BEATH, A., F. CHRISTIA, G. EGOROV, AND R. ENIKOLOPOV (2016): “Electoral Rules and Political Selection: Theory and Evidence from a Field Experiment in Afghanistan,” *The Review of Economic Studies*, 83, 932–968.
- BENOIT, K. (2000): “Which electoral formula is the most proportional? A new look with new evidence,” *Political Analysis*, 8, 381–388.
- BESLEY, T., J. G. MONTALVO, AND M. REYNAL-QUEROL (2011): “Do educated leaders matter?” *The Economic Journal*, 121, F205–227.
- BESLEY, T. AND T. PERSSON (2018): “Organizational dynamics: Culture, design, and performance,” *mimeo*.
- 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, 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, 167–181.
- 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.
- BORENSTEIN, S. (1991): “The dominant-firm advantage in multiproduct industries: Evidence from the US airlines,” *Quarterly Journal of Economics*, 106, 1237–1266.
- BROLLO, F., T. NANNICINI, R. PEROTTI, AND G. TABELLINI (2013): “The political resource curse,” *American Economic Review*, 103, 1759–1796.
- BUISSERET, P. E. AND C. PRATO (2017): “Electoral Accountability in Multi-Member Districts,” *mimeo*.
- 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, 767–779.

- CALONICO, S., M. D. CATTANEO, M. H. FARRELL, AND R. TITIUNIK (2016): “rdrobust: Software for regression discontinuity designs,” *Stata Journal*, 17, 372–404.
- CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): “Robust nonparametric confidence intervals for regression-discontinuity designs,” *Econometrica*, 82, 2295–2326.
- CAREY, J. M. (2008): *Legislative voting and accountability*, Cambridge University Press.
- CAREY, J. M. AND S. HIX (2011): “The electoral sweet spot: Low-magnitude proportional electoral systems,” *American Journal of Political Science*, 55, 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, 417–439.
- CARROLL, R. A. AND M. NALEPA (2018): “Electoral systems and programmatic parties: The institutional underpinnings of parties’ ideological cohesion,” *mimeo*.
- CATTANEO, M. D., B. R. FRANDBEN, AND R. TITIUNIK (2015): “Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate,” *Journal of Causal Inference*, 3, 1–24.
- CATTANEO, M. D., L. KEELE, R. TITIUNIK, AND G. VAZQUEZ-BARE (2016): “Interpreting regression discontinuity designs with multiple cutoffs,” *Journal of Politics*, 78, 1229–1248.
- CHATTOPADHYAY, R. AND E. DUFLO (2004): “Women as policy makers: Evidence from a randomized policy experiment in India,” *Econometrica*, 72, 1409–1443.
- CIRONE, A. E. (2017): “Dual Mandates, Patronage, and Partisanship: Evidence from a Regression Discontinuity Design in France,” *mimeo*.
- COX, G. W. (1990): “Centripetal and centrifugal incentives in electoral systems,” *American Journal of Political Science*, 34, 903–935.
- COX, G. W., J. H. FIVA, AND D. M. SMITH (2016): “The contraction effect: How proportional representation affects mobilization and turnout,” *Journal of Politics*, 78, 1249–1263.
- (2019): “Parties, Legislators, and the Origins of Proportional Representation,” *Comparative Political Studies*, 52, 102–133.
- DAL BÒ, E. AND F. FINAN (2018): “Progress and Perspectives in the Study of Political Selection,” *Annual Review of Economics*, 10, 541–575.
- DAL BÒ, E., F. FINAN, O. FOLKE, T. PERSSON, AND J. RICKNE (2017): “Who Becomes A Politician?” *Quarterly Journal of Economics*, 132, 1877–1914.
- DEBOWIZC, D., A. SAPORITI, AND Y. WANG (2016): “Redistributive Politics, Power Sharing and Fairness,” *Luxembourg Income Study Working Paper Series No 681*.
- DEBREU, G. (1952): “A social equilibrium existence theorem,” *Proceedings of the National Academy of Sciences*, 38, 886–893.

- DIERMEIER, D. AND T. J. FEDDERSEN (1998): “Cohesion in legislatures and the vote of confidence procedure,” *American Political Science Review*, 92, 611–621.
- 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, 200–212.
- EGGERS, A. C. (2015): “Proportionality and Turnout: Evidence From French Municipalities,” *Comparative Political Studies*, 48, 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*.
- FOLKE, O. AND J. RICKNE (2017): “The ballot rank hierarchy and the irrelevance of marginal seats: Party nomination strategies in closed and flexible PR,” *mimeo*.
- FUJIWARA, T. (2011): “A regression discontinuity test of strategic voting and Duverger’s law,” *Quarterly Journal of Political Science*, 6, 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, 369–398.
- GAGLIARDUCCI, S., T. NANNICINI, AND P. NATICCHIONIA (2011): “Electoral rules and politicians’ behavior: A micro test,” *American Economic Journal: Economic Policy*, 3, 144–174.
- GALASSO, V. AND T. NANNICINI (2011): “Competing on good politicians,” *American Political Science Review*, 105, 79–99.
- (2017): “Political selection under alternative electoral rules,” *Public Choice*, 171, 257–281.
- GALLAGHER, M. (1991): “Proportionality, disproportionality and electoral systems,” *Electoral Studies*, 10, 33–51.
- GROFMAN, B. (2008): “The Impact of Electoral Laws on Political Parties,” in *The Impact of Electoral Laws on Political Parties*, ed. by D. A. Wittman and B. R. Weingast, Oxford University Press.
- GROSSMAN, S. J. AND O. D. HART (1986): “The costs and benefits of ownership: A theory of vertical and lateral integration,” *Journal of Political Economy*, 94, 691–719.

- HERRERA, H., A. LLORENTE-SAGUER, AND J. C. MCMURRAY (2019): “The Marginal Voter’s Curse,” *Economic Journal*, forthcoming.
- HERRERA, H., M. MORELLI, AND S. NUNNARI (2016): “Turnout Across Democracies,” *American Journal of Political Science*, 60, 607–624.
- HERRERA, H., M. MORELLI, AND T. R. PALFREY (2014): “Turnout and power sharing,” *Economic Journal*, 124, F131–F162.
- HERRON, E. S., R. J. PEKKANEN, AND M. S. SHUGART (2018): *The Oxford Handbook of Electoral Systems*.
- 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, 68–81.
- (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, 829–853.
- JONES, B. F. AND B. A. OLKEN (2005): “Do leaders matter? National leadership and growth since World War II,” *Quarterly Journal of Economics*, 120, 835–864.
- KIRCHHEIMER, O. (1990): “The catch-all party,” *The West European Party System*, 50–60.
- KOPPEL, M. AND A. DISKIN (2009): “Measuring disproportionality, volatility and malapportionment: axiomatization and solutions,” *Social Choice and Welfare*, 33, 281–286.
- KREHBIEL, K. AND Z. PESKOWITZ (2015): “Legislative organization and ideal-point bias,” *Journal of Theoretical Politics*, 27, 673–703.
- KSELMAN, D. M. (2017): “Legislative Equilibria with Intra-Party Voting,” *mimeo*.
- LEE, D. S. AND T. LEMIEUX (2010): “Regression discontinuity designs in economics,” *Journal of Economic Literature*, 48, 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, 807–859.
- LEVINE, D. I. (1991): “Cohesiveness, productivity, and wage dispersion,” *Journal of Economic Behavior & Organization*, 15, 237–255.
- LIJPHART, A. (1995): *Electoral Systems and Party Systems*, Oxford University Press, USA.
- (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 (2016): “Are Voters Rational?” *VATT Working Papers* 50.
- MATAKOS, K., O. TROUMPOUNIS, AND D. XEFTERIS (2016): “Electoral rule disproportionality and platform polarization,” *American Journal of Political Science*, 60, 1026–1043.
- MCCRARY, J. (2008): “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142, 698–714.
- MERILÄINEN, J. (2018): “Politician Quality, Ideology and Fiscal Policy,” *mimeo*.
- MERILÄINEN, J. AND J. TUKIAINEN (2016): “Primary Effect in Open-List Elections,” *VATT Working Papers* 79.
- MILESI-FERRETTI, G. M., R. PEROTTI, AND M. ROSTAGNO (2002): “Electoral systems and public spending,” *Quarterly Journal of Economics*, 609–657.
- MORRIS, S. (2000): “Contagion,” *Review of Economic Studies*, 67, 57–78.
- PERSSON, T., G. ROLAND, AND G. TABELLINI (2007): “Electoral rules and government spending in parliamentary democracies,” *Quarterly Journal of Political Science*, 2, 155–188.
- PERSSON, T., G. TABELLINI, AND F. TREBBI (2003): “Electoral rules and corruption,” *Journal of the European Economic Association*, 1, 958–989.
- PERSSON, T. AND G. E. TABELLINI (2002): *Political economics: Explaining economic policy*, MIT press.
- (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, 269–278.
- REAGANS, R. AND B. MCEVILY (2003): “Network structure and knowledge transfer: The effects of cohesion and range,” *Administrative Science Quarterly*, 48, 240–267.
- REICH, B. (2016): “The diffusion of innovations in social networks,” *mimeo*.
- SANZ, C. (2017): “The Effect of Electoral Systems on Voter Turnout: Evidence from a Natural Experiment,” *Political Science Research and Methods*, 5, 689–710.
- SAPORITI, A. (2014): “Power sharing and electoral equilibrium,” *Economic Theory*, 55, 705–729.
- SAVOLAINEN, R. (2018): “Does Political Experience Influence Policy Positions?” *mimeo*.

- SEKHON, J. S. AND R. TITIUNIK (2017): “On interpreting the regression discontinuity design as a local experiment,” in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, Volume 38)*, ed. by M. D. Cattaneo and J. C. Escanciano, Emerald Publishing Limited, 1–28.
- 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, 818–820.
- TAAGEPERA, R. AND B. GROFMAN (2003): “Mapping the indices of seats–votes disproportionality and inter-election volatility,” *Party Politics*, 9, 659–677.
- TAAGEPERA, R. AND M. S. SHUGART (1989): *Seats and votes: The effects and determinants of electoral systems*, Yale University Press New Haven.
- THEIL, H. (1969): “The desired political entropy,” *American Political Science Review*, 63, 521–525.
- TSEBELIS, G. (2002): *Veto Players: How Political Institutions Work*, Princeton, NJ: Princeton University Press.
- WAGNER, M. AND O. RUUSUVIRTA (2012): “Matching voters to parties: Voting advice applications and models of party choice,” *Acta Politica*, 47, 400–422.
- WASHINGTON, E. L. (2008): “Female socialization: how daughters affect their legislator fathers,” *American Economic Review*, 98, 311–32.