Party Nomination Strategies in List Proportional Representation Systems *

Peter Buisseret† Olle Folke ‡ Carlo Prato§ Johanna Rickne¶

July 18, 2019

Abstract

In list proportional representation (PR) systems, parties critically shape political selection. Despite the ubiquity of these electoral rules, however, the incentives that guide the construction of party lists remain poorly understood. We propose a theory of party list choice and elections in list PR systems. Our results describe how a party allocates candidates of heterogenous quality across list ranks depending on (1) its electoral goals and (2) its competitive environment. We test our predictions on the universe of Swedish local politicians from 1982 to 2014. While parties assign better candidates to higher ranks at all ballot levels, the pattern is most pronounced among electorally advantaged parties, i.e., those with the strongest prospect of controlling the executive. These results contrast with existing accounts of candidate selection in both multi- and single-member electoral contexts, which emphasize that parties prioritize candidate quality in their nomination strategies only when constrained by electoral incentives. Our results, instead, suggest that the principal demand for high quality politicians derives from parties, rather than voters.

*Financial support from the Swedish Research Council and the Torsten Söderberg Foundation is gratefully acknowledged. We thank Chiara Superti, Giovanna Invernizzi, Daniel Smith, Thomas Däubler, Wiola Dziuda, Anthony Fowler, Shigeo Hirano, John Marshall, and Ariel White for valuable comments and suggestions. We also thank Jonas Ahlerup, Johan Arntyr, Sirus Dehdari, Roza Khoban, and Elin Molin for excellent research assistance.

†Harris School of Public Policy, University of Chicago, *Email*: pbuisseret@uchicago.edu

‡Department of Government, Uppsala University *Email*: olle.folke@statsvet.uu.se

§Department of Political Science, Columbia University, *Email*: cp2928@columbia.edu

¶Swedish Institute for Social Research, Stockholm University, *Email*: johanna.rickne@sofi.su.se

1. Introduction

The quality of politicians is a fundamental input to effective government. For this reason, scholars emphasize how elections empower citizens to select able policymakers (e.g., Fearon, 1999). In reality, however, political selection in most democracies is fundamentally influenced by political parties. By acting as gate-keepers to aspiring candidates, parties shape the choices available to voters and, ultimately, the quality of elected officials. In order to understand the determinants of political selection, we must therefore understand the forces that shape how parties select their candidates.

In this paper, we use both theory and data to study how parties structure candidate selection in multi-member districts under list proportional representation ("list PR") electoral systems. Approximately two-thirds of the world's legislative seats are contested under list PR—far in excess of those that are contested in single-member districts.[1] Yet party nomination strategies in list PR contexts remain poorly understood (Dal Bó and Finan, 2018).

PR systems are inherently complexity—especially relative to single-member district (SMD) elections. Under SMD, a party's ballot is an *individual* candidate. Under list PR, however, a party's ballot is a ranking of *multiple* candidates. This generates a richer set of nomination strategies for political parties. And, it makes it harder to conceptualize voting behavior, since the link between a vote for a party list or a candidate and election outcomes is much less clear-cut than in single-member systems.

Our goal is to understand how parties' electoral objectives shape the relationship between a candidate's quality and her list assignment. In what contexts should we expect parties to promote higher quality candidates to more electorally secure ranks? Do elections incentivize parties to promote higher quality candidates? And, what can we learn about parties' goals and priorities from real-world electoral lists?

Our Approach. We first introduce a new theoretical framework with three ingredients: an individual calculus of voting, strategic list choice by parties, and an election. Each party has a pool of candidates that are differentiated by their quality, or *valence*: a set of observable characteristics that voters and parties value. Each party orders its candidates in a list; all else equal, a higher position on the list implies a higher prospect of election. Voters observe each party's list and cast their vote in favor of one of a party (or, in flexible list settings, a candidate within a party's list). Parties win seats in proportion to their vote shares. Under closed lists, seats are filled by candidates in the order that they are ranked. Under flexible lists, a

[1] According to the *Database of Political Institutions* (Cruz, Keefer and Scartascini, 2016), as of 2015 legislative elections in 94 out of 147 democracies employ proportional representation.

candidate with enough preference votes is elected regardless of her rank.

Our calculus of voting assumes that voters have partisan preferences, but also prefer more competent candidates.[2] We derive a voter's induced preferences over party lists as a function of how parties rank their candidates, the voter's partisan preferences, and her forecast of each party's popularity and thus its likely share of seats.

We then introduce strategic party selectorates. The selectorate could represent a local party elite, or a group of factions or individuals. We assume that the selectorate values three things: winning seats, wielding executive authority (e.g., appointing the prime minister or mayor), and securing the election of high-quality candidates that improve their post-election legislative success. Our framework allows us to consider different ways in which selectorates might prioritize these considerations—to the extent that they do not perfectly align.

To understand why these goals might not align, suppose that parties are concerned with winning seats—considered by Engstrom (1991) to be "the ultimate goal of most [party] selectors." If voters favor high-quality candidates, we might expect that electoral incentives would encourage party leaders to put better candidates in better ranks. Surprisingly, we show that this conjecture is false. To see why, notice that under PR, each party is near-certain to secure the election of its highest-ranked candidates. A voter therefore recognizes that her vote is likely to be inconsequential for the election prospects of candidates located within these safe ranks. Instead, her vote is more likely to matter for the prospects of candidates in electorally *marginal* ranks, further down a party's list. A party therefore raises its electoral appeal by placing better candidates in these marginal, ranks where voters expect their choices to matter the most. This is the essence of our first prediction, which we call the *Marginal Rank Hypothesis*.

In practice, of course, parties do *not* care solely about winning seats—parties also value the election of their best talent. Acknowledging the reality that parties care about electing competent legislators, we derive a competing prediction, the *Top Down Rank Hypothesis*. It provides sufficient conditions under which (1) parties place better candidates in safer ranks, even outside of the leadership roles in the very highest ranks, and (2) that the incentive to do so is especially strong amongst electorally advantaged parties—those that are relatively likely to control executive authority due to a partisan or popularity advantage.

Our second hypothesis contrasts with many prominent explanations for candidate selection both in PR and SMD contexts. A classic theme of existing work is that voters value high quality candidates, but political parties balance a range of other considerations when selecting

[2] There is significant evidence that voters do care about the characteristics of individual candidates in list PR contexts; see, for example, Shugart, Valdini and Suominen (2005).

candidates. These considerations include: loyalty (Galasso and Nannicini, 2015), ideological congruence (Serra, 2011), or geographic balance (Shugart *et al.*, 2005), to name a few. This logic is rife in the literature on primary adoption. Most recently, for example, Slough, York and Ting, 2018 predict that electorally disadvantaged parties are most likely to favor primaries because their party elites cannot otherwise commit to selecting high-quality candidates that make the party's ballot more electorally attractive, but with whom elites need to share power.

If, in our setting, the demand for high quality politicians arises principally from voters, rather than parties, one would expect that advantaged parties—whose electoral constraints are relatively weakest—would be the most likely to deviate from the principal of placing better candidates in safer ranks. Our Top Down Rank Order Hypothesis predicts precisely the *opposite*, arguing that it is the parties with the *most* likely prospect of controlling executive authority that are most likely to assign better candidates to safer ranks.

Our theoretical framework applies to any parliamentary context—local or national—in which (i) voters select politicians to represent them in multi-member districts under closed or flexible (i.e., not completely open) list PR and (ii) candidate selection decisions are not completely decentralized, so that the notion of a party's selectorate is analytically meaningful.

In order to bring our predictions to the data, we would ideally want (1) a large number of real-world party lists, (2) precise measures of individual candidate quality for each candidate on each list, and (3) a methodology to forecast the expected seat shares of each party in each district election. Measures of candidate quality would allow us to identify the distribution of valence across a party's list, and forecasting methodology would allow us to identify which party's seats are safe versus marginal (to evaluate the Marginal Rank hypothesis), and also to identify which parties are relatively likely to wield executive authority (to evaluate the Top Down Rank Order hypothesis).

We make use of a comprehensive dataset that responds to the above desiderata, from Sweden. Our empirical analysis focuses on municipal elections, where the local branches of the eight established political parties operate in nearly every one of the 290 municipalities. Each municipality operates as a classic parliamentary democracy, with elections to the council (i.e., legislative assembly) and subsequent appointment of the mayor (i.e., the chief executive) by the largest party in the governing coalition. The parties that have representation in the parliament also have local party branches across all, or most, municipalities and these local parties enjoy complete autonomy in determining the rank-order of their electoral ballots (we describe the process in Section 4).

Our analysis leverages administrative data on the universe of nominated and elected

Swedish politicians between 1991 and 2014. These data are high-quality, never self-reported, and contain very few missing values. We use the information in these registers to develop *four distinct measures* of candidate quality, based on leadership skills, cognitive abilities, labor market outcomes, and education.

In each of seven elections we generate the complete quality-rank ordering of candidates, as well as measures of which ranks are electorally marginal, for over 6,000 local parties. With this level of statistical power, we can comprehensively evaluate our predictions, even after restricting to subsets of the data where our theoretical model predicts the clearest empirical patterns.

We decisively reject the marginal rank hypothesis and find strong support for the top-down rank order hypothesis. In particular, parties with a history of executive control also display the strongest (positive) relationship between ballot rank and quality, even excluding top ranks that are traditionally reserved to leadership positions. The fact that virtually all parties place their best candidates in the top (uncontested) ranks further invalidates the marginal rank hypothesis. But the fact that it is the advantaged parties—that face the weakest electoral constraints—that are most likely to place better candidates in higher ranks at all ballot levels, suggests that it is the parties that drive the demand for high quality politicians.

Contribution. Since our empirical setting reflects a classical parliamentary democracy, the significance of our findings extends beyond the Swedish context.

We bring new theory and evidence to bear on the question of how electoral systems affect intra-party organization and candidate selection. Scholars have devoted considerable effort to understanding candidate selection in single-member electoral contexts, where "the primary is [not] a method of nomination; it is the election" (Key, 1964). The choice of candidate selection process reflects competing objectives within a party's selectorate—for example, obtaining better quality candidates through more competition at the nomination stage (primaries) versus obtaining candidates that are more aligned with party elites (closed) (Serra, 2011, Snyder and Ting, 2011, or Slough *et al.*, 2018). In line with this perspective, we also generate predictions about candidate nomination outcomes as the resolution of competing objectives within a party's selectorate, but with a focus on list PR and multi-member districts.

As such, this paper contributes to a growing literature on political selection in list PR systems. Existing contributions have focused on more descriptive aspects of selection, such as female representation on party lists (Baltrunaite, Bello, Casarico and Profeta, 2014; Esteve-Volart and Bagues, 2012; Besley, Folke, Persson and Rickne, 2017) and the presence of candidates with local ties and experience (Shugart *et al.*, 2005). Conversely, this paper focuses on competence and quality—attributes that should be universally valued by voters and party leaders alike.

Contrary to a large body of work on the effects of electoral rules on *inter*-party competition—e.g., the number of parties (Duverger, 1959), our focus is on their consequences for *intra*-party organization. Scholars have acknowledged the importance of party's legislative hierarchy for electoral outcomes. Unique to our study is the focus on parties' ranks in list PR: in single-member contexts (e.g., Weingast and Marshall, 1988), only one candidate appears on the ballot for each party—making the electoral hierarchy degenerate in any single race in any given district; in open-list settings (e.g., Patty, Schibber, Penn and Crisp, Forthcoming), the party's list ranking plays no role in determining the order in which candidates are elected. Our work therefore joins a small literature that conceptualizes lists as an expression of a party's internal hierarchy in (closed and flexible—or semi-closed) list PR, and studies how these hierarchies are shaped by a party's electoral environment (Hobolt and Høyland, 2011; Nemoto and Shugart, 2013; André, Depauw, Shugart and Chytilek, 2015; Crutzen and Sahuguet, 2017).

A few closely related papers show how party lists in preferential systems can provide incentives for effort provision (Crutzen, Flamand and Sahuguet, 2017) or act as internal contests for promotion (Meriläinen and Tukiainen, 2018; Folke, Persson and Rickne, 2016). Conversely, our theory conceptualizes candidate quality as a fixed attribute and identifies a set of forces that shape political selection irrespective of list type.

Closest to our work is Matakos, Savolainen, Troumpounis, Tukiainen and Xefteris (2018), who also identify a trade-off between a party's electoral support and its post-election success in legislative politics. In their setting, the trade-off turns on the ideological composition of the list, i.e., cohesion versus breadth of electoral appeal, in contrast with our focus on candidate quality.

Organization. We develop our theoretical model of elections and legislative policymaking in Section 2. We solve the model and obtain theoretical predictions in Section 3. Background information on our empirical setting is provided in Section 4. Section 5 describes our data, and develops measures of candidate quality and party competition. We use these to test our theoretical predictions in Section 6. A discussion of alternative frameworks follows, and a conclusion. Proofs, additional empirical results, sensitivity tests and robustness checks, as well as theoretical results flexible list systems are contained in a Supplemental Appendix.

2. Model

To streamline presentation, our benchmark model focuses on a *closed list* PR setting. In Supplemental Appendix C, we extend our framework to flexible list settings. The theoretical predictions in Propositions 1 and 2 apply to both systems.

Agents. We consider an interaction between a unit mass of voters, and two local political parties, A and B. The parties compete in an election for control of a legislative assembly, which consists of five elected representatives. The assembly represents a national parliament, or a municipal council. Each party fields five candidates that differ in their observed quality (or, human capital) q; without loss of generality we denote candidates by their quality level: $q \in Q \equiv \{1, 2, 3, 4, 5\}$, where $q = 1$ denotes the lowest level, and $q = 5$ denotes the highest level.

The interaction proceeds in two stages: an *election* and a subsequent *legislative interaction*.

Election. Each party simultaneously chooses a *list* that determines the order in which its candidates appear on the ballot. Specifically, a list for party $J \in \{A, B\}$ is $l_J \in L(Q)$—where $L(Q)$ is the set of strict orderings over Q—and its k^{th} element is denoted $l_J(k)$.

Our baseline model focuses on *closed list* contexts. Thus, after the parties choose their lists, voters cast their ballots in an election for one of the two party lists.[3] After votes are cast, each party is allocated n_J seats, where $n_A + n_B = 5$. Seats are allocated proportionally across the parties: for each seat $k \in \{1, ..., 5\}$ there is a threshold vote share $\pi(k) \in (0, 1)$ that a party must surmount in order to secure that seat. So, when party J's vote share π_J is between $\pi(k)$ and $\pi(k+1)$, $n_J = k$, and $n_{-J} = 5 - k$. In keeping with real-world PR systems, we impose $\pi(3) = \frac{1}{2}$ and for all $k < 3$, $\pi(k) = 1 - \pi(6 - k)$.

After the election, each party allocates its seats to the first n_J candidates that appear on its list, denoted $S_J \equiv \{l_J(1), \ldots, l_J(n_J)\}$. Thus, the assembly is $S \equiv S_A \cup S_B$. The assembly is headed by a chairperson, the *chief executive*, denoted $m \in S$. In the local context the chief executive is the mayor, but in the context of a national election it could represent the prime minister. For simplicity, we assume that the party obtaining a majority of the seats appoints the chief executive.[4]

Voters care about the quality of governance. This depends positively on the total human capital of the assembly—$\sum_{i \in S} i \equiv \overline{q}$—and the human capital of the chief executive (m). These preferences are represented by the payoff function

$$H(\overline{q}, m) : \{9, ..., 21\} \times Q \to \mathbb{R}.$$

[3] By focusing on candidate selection, we implicitly assume that local parties have limited capacity to adjust voters' perceptions of their policy platforms and other programmatic aspects in a given election cycle. We view this assumption as natural in our empirical setting.

[4] The model can be extended without any significant additional insights to multiple executive offices and to a probabilistic allocation of the chief executive across parties.

We impose minimal structure by assuming only that governance quality increases in the aggregate human capital of the assembly and the human capital of the chief executive.

Assumption 1. *H strictly increases in each of its arguments.*

In other words, voters want the best possible candidates to sit in the legislative assembly. Notice that voters may care relatively more about the quality of mayor than that total quality of rank-and-file legislators. This would capture a policy-making context in which mayors enjoy considerable agenda-setting power, or executive discretion. In municipal contexts, for instance, the mayor may be the only paid or full-time political position.

Legislative Interaction. After the election, political parties divide a *surplus* with total normalized to one. This surplus captures both material resources that can be channeled to specific individuals and groups (e.g., jobs) as well as policymaking authority—including the ability to allocate funds to the party's preferred public goods. We assume that party $J \in \{A, B\}$'s share of the surplus depends on (i) the number of its allocated seats in the assembly (n_J), (ii) the human capital of the chief executive, and (iii) its share $\overline{q}_J(S) \equiv \frac{\sum_{i \in S_J} i}{\overline{q}}$ of the aggregate human capital of the assembly. A party that obtains k seats in assembly S enjoys a payoff

$$G(k, m, \overline{q}_J) : \{1, \ldots, 5\} \times Q \times [0, 1] \to [0, 1].$$

Assumption 2. *$G(k, m, \overline{q}_J)$ (i) strictly increases in k, (ii) weakly increases in $\overline{q}_J$, and (iii) weakly increases in m if $k \geq 3$ and weakly decreases in m if $k < 3$.*

Assumption 2 states that a party's share of the surplus always increases with its share of the seats. The party's share may increase continuously or even discontinuously at critical thresholds such as a majority. Furthermore, a party's value *increases* with the quality of the chief executive when it appoints the chief executive from its own ranks and does *not* increase when the opposing party appoints a higher-quality chief executive; and, it does *not* decrease when its share of the total skills and experience in the assembly increases relative to the opposing party's share.

In a highly majoritarian legislative context where all power derives from majority control, we have $G(3, \cdot, \cdot) - G(2, \cdot, \cdot) \approx 1$. Nonetheless, our formulation also allows for a more progressive distribution of political power as a function of the number of seats. This could reflect an institutional context in which the distribution of political power more closely reflects the relative share of seats controlled by each party.

In Sweden and in other real-world settings, nomination decisions involve negotiations between multiple individuals, groups and factions. Our formulation does not assume that nomination processes are decided unilaterally by a single party leader; rather, we interpret the payoff function $G(\cdot,\cdot,\cdot)$ as a reduced-form description of a party's selectorate and the relative bargaining power of its various groups.

Voting Behavior. According to a process that we set out in detail, below, each voter i computes an equilibrium expected value of voting for each party V_J and votes for party B if and only if

$$V_B \geq V_A + \sigma_i + \xi \tag{1}$$

where σ_i is a voter-specific preference shock drawn from a zero-mean uniform with density ϕ and support $[-(2\phi)^{-1}, (2\phi)^{-1}]$, and ξ is an aggregate preference shock drawn from a distribution $F(x)$ with support $[-(2\psi)^{-1}, (2\psi)^{-1}]$ and density $f(\cdot) = \psi(1-\theta) + \theta\tilde{f}(\cdot)$. We assume that $\tilde{f}(\cdot)$ is single-peaked, continuous, differentiable and symmetrically distributed around its peak, $\overline{\xi}$. Without loss of generality, we assume that $\overline{\xi} > 0$. This implies $\mathbb{E}[\xi] > 0$, so that the aggregate shock is expected to favor party A. Higher values of $\overline{\xi}$ reflect a greater *electoral advantage* to party A: all else equal, party A expects to receive a larger share of seats than party B. Higher values of ψ reflect voter preferences that more responsive to variation in the quality of a party's candidates in the election, as opposed to other factors such as partisanship.

Thus, our framework does not assume that voters' behavior and attention are *solely* driven by candidate quality: it may also be guided by programmatic and expressive considerations. What is crucial is that *some* voters respond to changes in the composition of party lists—a premise for which we provide evidence in Section 7.

Reflecting real-world electoral contexts, we allow for the possibility that each party is sure to win at least one seat:

Assumption 3. *Each party is assured of winning at least one seat:*

$$\underline{\pi} = \frac{1}{2} + \phi(-\overline{V} + \underline{V} - (2\psi)^{-1}) \in (\pi(1), \pi(2)),$$

where $\overline{V}$ and $\underline{V}$ denote the largest and smallest values from a party's list.

Assumption 3 implies that each party expects to receives at least one seat (an *uncontested* seat). We refer to the remaining three seats as *contested* seats.

Summary of Timing. To summarize, the timing of the events is as follows:

1. Each party constructs an electoral ballot l_J.

2. Each voter computes values V_A and V_B.

3. The common (ξ) and idiosyncratic (σ_i) shocks are realized and votes are cast.

4. Seats are allocated and the majority-winning party appoints the chief executive.

Discussion. Our framework extends easily to an arbitrary district magnitude. As we detail below, however, our empirical approach groups each party's list ranks into categories: our hypotheses about the quality of a candidate in single rank can be interpreted as the average quality of candidates in the corresponding rank category in our data.

Our model focuses on a *closed* list system, in which voters cannot override the order in which candidates on the party's electoral ballot are elected. In Supplemental Appendix C, we extend our model to flexible list settings. The reason is that our election data contains both periods of *closed* (pre-1998) and *flexible* (post-1998) list PR. Our Appendix shows that all of our theoretical predictions extend to flexible lists, and thus apply to the complete sample of our data.

3. Theoretical Results

Chief Executive Appointment. Recall that whichever party wins a majority of seats appoints the chief executive. Since a party benefits from higher quality leadership when it enjoys majority status, it chooses its best elected politician to lead the assembly.

Lemma 1. *If $G(k, m, \overline{q})$ strictly increases in m, then the majority party appoints its highest-quality elected official to be chief executive.*

We may therefore denote the equilibrium identity of the chief executive as $\hat{m}(n_A, l_A, l_B)$, which depends on the lists l_A and l_B, as well as the number of seats n_A awarded to party A (recall that $n_B = 5 - n_A$):

$$\hat{m}(k, l_A, l_B) = \max\{l_A(1), \ldots, l_A(k)\}\mathbf{I}[k \geq 3] + \max\{l_B(1), \ldots, l_B(5-k)\}\mathbf{I}[k < 3]. \qquad (2)$$

Induced Voter Preferences. Given the party's lists, l_A and l_B, voters decide whether to cast a vote for party A's list, or party B's list. A voter's choice takes into account her conjectures about (i) the anticipated quality of the post-election assembly—including parties' appointment decisions—and (ii) the impact that her vote will have on the outcome of the election.[5]

[5] Since no voter can be decisive for the outcome, *any* voting behavior is consistent with optimality. The behavior that we develop below is a heuristic that we believe is most faithful to the restriction to weak dominance in the single-member context.

If there were only a single representative (i.e., a single-member system) in a two-candidate contest, the election outcome turns on a single event: one candidate winning a majority of the votes. Thus, in the hypothetical event that a voter *could* have been decisive for the outcome, she would always prefer to cast her ballot in favor of the candidate she most prefers. This motivates a restriction to *sincere* voting in single-member contexts.

In multi-member districts, by contrast, the consequences of a vote in favor of either ticket depend on how voters in the same district cast their ballots. To see this, consider the following example.

Rank	l_A	l_B
1	Theseus	Titania
2	Hermia	Oberon
3	Lysander	Cobweb
4	Helena	Puck
5	Demetrius	Moth

Suppose that a voter is in a district where party A is very popular—to such an extent that the party is certain to win at least three seats. This means that Theseus, Hermia and Lysander are certain to be elected from A's list. Conversely: Cobweb, Puck and Moth are certain *not* to be elected from party B's list. Recalling that each party has a single safe seat, the only uncertainty is therefore whether A will win only three seats, or whether instead the party will win four seats. A vote for A's list is most likely to determine whether its *fourth*-ranked candidate, Helena, is elected, rather than B's *second*-ranked candidate, Oberon.

Alternatively, the voter may reside in a more competitive district, and anticipate that other voters will be evenly split across parties A and B. In this district, the first-order uncertainty is whether party A wins two or three seats. A vote for A's list is most likely to determine whether its *third*-ranked candidate, Lysander, is elected, rather than B's *third*-ranked candidate, Cobweb.

Finally, the voter may expect that A will win very little support, so that her main uncertainty is whether A wins one or two seats. A vote for A's list is most likely to determine whether its *second*-ranked candidate, Hermia, is elected, rather than B's *fourth*-ranked candidate, Puck.

More generally, let $\tilde{H}(k, l_A, l_B)$ denote voters' payoff as a function of the number of seats awarded to A, the lists, and given $\hat{m}(k, l_A, l_B)$:

$$\tilde{H}(k, l_A, l_B) = H\left(\sum_{i=1}^{k} l_A(i) + \sum_{i=1}^{5-k} l_B(i), \hat{m}(k, l_A, l_B)\right).$$

If a voter were relatively sure that her vote would determine whether A wins k seats or $k-1$ seats, her relative value of voting for A's list would be $\tilde{H}(k, l_A, l_B) - \tilde{H}(k-1, l_A, l_B)$. But how does a voter assess the relative prospect that her vote matters for the election of each of A's ranks $k = 1, ..., 5$? We develop a voting calculus that answers this question.

Lemma 2. *For any party lists l_A and l_B, there exists a pair $V_A(l_A, l_B)$ and $V_B(l_A, l_B)$, and attention weights $\tau(k) \in (0, 1)$ for $k \in \{2, 3, 4\}$ satisfying $\sum_{k=2}^{4} \tau(k) = 1$, such that an instrumental voter's value from party A is:*

$$V_A(l_A, l_B) = \sum_{k=2}^{4} \tau(k)\ \tilde{H}(k, l_A, l_B),$$

while her value from party B is:

$$V_B(l_A, l_B) = \sum_{k=2}^{4} \tau(k)\ \tilde{H}(k-1, l_A, l_B).$$

Moreover, when θ is small enough, for each pair of lists (l_A, l_B), the difference $\Delta(l_A, l_B) = V_A(l_A, l_B) - V_B(l_A, l_B)$ is unique.

Lemma 2 highlights how voters evaluate each party's list, not only as a function of the candidates and their order on the ballot, but also how they conjecture that other voters will behave. Each *attention weight*, $\tau(k)$, is the probability that a voter is decisive for the k^{th}-ranked candidate of party A, *conditional on being decisive*. The weight $\tau(k)$ describes the fraction of his attention that is devoted to comparing the relative value from electing the k^{th}-ranked party A candidate, versus the $(6-k)^{th}$ party B candidate, i.e., raising party A's control of the assembly from $k-1$ to k seats.

Because each party always elects at least one candidate, the top-ranked candidate on each list is certain to be elected. Conversely, the bottom-ranked candidate on each list is certain *not* to be elected. Thus, voters expect to be decisive only for candidates within the *contested* ranks, i.e., the second, third and fourth ranks.[6]

[6] Nonetheless, a party's top-ranked candidate may still play an important role. If, for example, a party's

Our benchmark presentation focuses on closed lists, where voters anticipate the seats will be filled in the order that candidates are ranked. In a flexible list setting, however, each voter must also account for the possibility that if a sufficient number of voters cast their preference votes for one of the candidates, that candidate will be elected regardless of where she is ranked. This does not alter the logic of Lemma 2, however, and we show that it extends to flexible lists in Supplemental Appendix C. Henceforth, we assume that θ is small enough that the Lemma holds.

Party Nomination Strategies. Our two main results characterize equilibrium party lists. The results are obtained under alternative perspectives on (i) how voters care about different aspects of the aggregate quality of the assembly, i.e., the function H, and (ii) aspects of a party's elected members that are most effective for the pursuit of legislative goals, e.g., majority control or maximizing the talent of its elected representatives. These aspects are determined by the function G.

Marginal Rank Hypothesis. We initially consider a stark perspective: that what parties consider most important to pursuing legislative goals is winning as many seats as possible.

Definition 1. Parties are *seat-motivated* if $G(k, q, \overline{q})$ depends only on the number of seats, k, i.e., it is constant in q and $\overline{q}$.

We treat this objective as a benchmark, reflecting that a principal goal of parties is to secure the election of their candidates (e.g., Downs, 1957): it states that the only way in which parties value electoral outcomes is through the number of seats that they are allotted in the council.[7]

In order to explicitly connect our predictions to our subsequent empirical results, we refer to the second rank on a each party's list in our theoretical framework as its *advantaged* competitive rank, and the fourth rank as its *disadvantaged* competitive rank.

Proposition 1. *(Marginal Rank Hypothesis) When parties are purely seat-motivated, within the contested ranks:*

1. *A's best candidates are not located in its* advantaged *ranks:* $l_A(2) < \max\{l_A(3), l_A(4)\}$.
2. *B's best candidates are not located in its* disadvantaged *ranks:* $l_B(4) < \max\{l_B(2), l_B(3)\}$.

Finally, when A's advantage, i.e., $\overline{\xi}$, is large enough:

$$l_A(2) < l_A(3) < l_A(4) \;\; \textit{and} \;\; l_B(2) > l_B(3) > l_B(4).$$

top-ranked candidate is also its highest quality candidate, a voter recognizes that when A wins the third seat the party also gets to appoint this top-ranked candidate to be the mayor.

[7] For example, when $G(3) - G(2) \approx 1$, parties almost exclusively care about securing a majority of the assembly

Proposition 1 establishes a direct link between the electoral salience of a list rank and a seat-motivated party's incentive to assign to it a high quality candidate. Since each voter voter places less attention on party A's second rank—correctly anticipating that her vote is relatively more consequential for party A's third or fourth seat, A's selectorate maximizes the list's electoral appeal by placing a relatively better candidate in either the third or fourth rank than in the second rank. For the same reason, party B does not place its best candidate in its fourth rank, i.e., its disadvantaged competitive rank: voters pay more attention to its second and third rank, which are associated with more competitive contests.[8] The final part of the proposition provides even stronger results, for a setting in which party A is very likely to win a large share of seats.

Top-Down Rank Order Hypothesis. An alternative view of parties and legislative institutions, however, places much greater emphasis on the role of individual politicians—in particular, legislative *leaders*—rather than seats, alone, as the route to legislative accomplishment. We now consider this perspective, and ask how it affects the organization of party lists.

Recall that a party's payoff $G(k, m, \overline{q}_J)$ depends on k, the number of seats it wins, m, the quality of the chief executive, and $\overline{q}_J$, its share of the legislative human capital. Recall our assumption that G weakly increases in k and in $\overline{q}_J$. We derived the marginal rank hypothesis under an especially stark form of this assumption: purely seat-motivated parties care only about their share of seats (k).

We consider an alternative view of how parties achieve their legislative goals—not merely by seats alone, but also by virtue of having skilled and able legislators, by assuming that G strictly increases in *both* $\overline{q}$ *and* k. This implies that—all else equal—a party is strictly more likely to accomplish its legislative goals when its share of the total human capital in the assembly increases.

We formalize the perspective that legislative leaders matter via two conditions: our first condition focuses on how *voters* view the determinants of policy outcomes.

Condition 1. *There exists κ small enough such that $H_1(\cdot, 5) < \kappa$, where H_1 denotes the derivative of H with respect to its first argument.*

Suppose that the assembly leadership—the chief executive—is the highest quality candidate (i.e., with quality 5) from either of the two parties. Condition 1 states that—given the identity of the assembly's leadership—the incremental value of more able rank-and-file politicians is not too large. Alternatively stated: voters recognize that high quality leaders are the

[8] The intuition is similar to political donors in a single-member district election focusing their efforts on relatively competitive districts.

primary drivers of good policy outcomes. This condition is very likely to hold in parliamentary contexts, where agenda-setting powers typically reside with the executive, from which the vast majority of legislative initiatives originate.

While Condition 1 is sufficient to generate a testable restriction on the data, we consider a related condition that describes how—from the perspective of parties—the skills of a legislative team interact with a party's agenda-setting authority to facilitate the achievement of legislative goals. The condition states that variation in the quality of a party's legislative team matters relatively more when the party holds a majority of seats. Consider the strictly decreasing list l^*:

$$l_J^*(1) = 5,\ l_J^*(2) = 4,\ l_J^*(3) = 3,\ l_J^*(4) = 2,\ l_J^*(5) = 1.$$

This list maximizes a party's average share of the assembly's human capital, for *every* realized allocation of seats between the parties.

The next condition says that the marginal value of an improvement in the quality of an elected representative is highest for a party when it holds majority status. In other words: institutional authority and human capital are *complements* to a party's legislative goals.

Condition 2. *For any $k \geq 3 > k'$, and any list $l' \neq l^*$:*

$$\tilde{G}(k, l^*, l^*) - \tilde{G}(k, l', l^*) > \tilde{G}(k', l^*, l^*) - \tilde{G}(k', l', l^*). \tag{3}$$

We emphasize that this condition is not needed for our second main result, but it provides a valuable comparative static that generates additional empirical implications, and thus more exacting tests on our data. Conditions 1 and 2 have a common theme: they emphasize that policy outcomes and goals are not accomplished by seats, alone; rather, the political dividend of majority control largely hinges on the skill and experience of legislative leaders.

The expected payoff to party J when its list is l_J and the opposing party $-J$ offers a list l_{-J} is:

$$\Pi_J(l_j, l_{-j}) = \sum_{k=1}^{4} \Pr(\pi(k) < \pi_J < \pi(k+1) | l_J, l_{-J}) \tilde{G}(k; l_J, l_{-J}). \tag{4}$$

Our second result establishes that when Condition 1 holds, the list l^* is a best response to *every possible list* that could be selected from the opposing party. If, in addition, Condition 2 holds, the incentive to promote higher quality candidates to higher ranks is always stronger for party A, the advantaged party. In other words: the party with the greatest relative value from promoting better candidates is the party that is most likely to win majority control of the assembly.

Proposition 2. *(Top-Down Rank Order Hypothesis) Suppose that G strictly increases in $\overline{q}$ and k.*

[1.] If Condition 1 holds, the list l^* is the unique best response to any opposing party's list.

[2.] If Condition 2 also holds, then for any list $l \neq l^*$:

$$\Pi_A(l^*, l^*) - \Pi_A(l', l^*) > \Pi_B(l^*, l^*) - \Pi_B(l', l^*).$$

Propositions 1 and 2 generate empirically testable predictions that we test using Swedish register data on the competence characteristics of the complete universe of politicians in the period 1982-2014.

4. Empirical Context

Sweden has three levels of political representation: the parliament, 21 counties and 290 municipalities. All three levels are concurrently elected, and voter turnout is usually between 80 and 90 percent. Local elections are politically and economically important: municipalities have significant political autonomy and control large shares of expenditures and employment sectors in the Swedish welfare state. Notably, they spend between 15 and 20 percent of GDP, and employ roughly 20 percent of the country's labor force. Election to a municipal council also serves as the principal route into national politics (Lundqvist, 2013; Dal Bó, Finan, Folke, Persson and Rickne, 2017).

Municipal councils vary between 21 and 101 members, with an average of 46. Representation is not subject to an explicit electoral threshold. A municipality is effectively a parliamentary democracy, in which voters only elect the members of the legislative assembly, and all executive appointments are made after the election. Nearly all executive positions are appointed by the governing majority, i.e. the political coalition or single political party that receives more than 50% of the council seats.

The chief executive is the municipal council board chair.For simplicity we henceforth refer to the office as the *mayor*. The mayor is always appointed by the largest party in the ruling coalition. It is the most powerful position in municipal politics. With the exception of large municipalities, the mayor is also the only political office that offers a full-time salary, as opposed to piece-rate compensation for meetings (Montin, 2015).

The Swedish party system is very stable, and most political parties have local branches in all municipalities. A strong norm in Swedish parties protects local branch autonomy in composing electoral ballots and determining post-election appointments. The only rules are that candidates must be above 18 years old and be residents of the municipality in which they stand for election. Some parties also employ gender quotas that are dictated by the central party organizations. Purely local parties, i.e. without representation in the national parliament, account for less than four percent of the party-election observations in our data.

The precise procedure for selecting and ranking candidates varies across parties. Parties on the ideological left mainly use internal nominations.[9] In this case, nominations for the ballot are made by the party's clubs and are then aggregated by an election committee. Clubs are organized around neighborhoods in the municipality and around factions such as the *Women's League* or the *Youth League*. The election committee consists of senior politicians—either former or current—who can make adjustments to the final list before it receives final approval at a member meeting.

Parties on the ideological center-right are more likely to use primaries to determine the rank-order of their ballot.[10] Members vote for their preferred candidates, and votes are aggregated to determine the rank-order on the ballot. Center-right parties are also organized in party branches and geographic clubs, and members loosely belong to factions that coordinate votes around their political interests.

Regardless of whether parties use internal nominations or primaries, however, local party leaders retain significant influence over the nomination process. Leaders are able to exert direct and indirect influence over the selection committee that administers the initial round of candidate selection and determines the list ranking. The party elite's influence is reflected in the fact that few changes are ever made to party lists by the party rank-and-file at members' meetings. Recall that our framework represents selectorate goals via the function $G(\cdot,\cdot,\cdot)$. Our restrictions—embodied in Assumption 2—are flexible enough to represent a range of elite and rank-and-file objectives, and how these perspectives are ultimately reconciled in nomination processes. In particular, since our main theoretical results are based on contrasting views of party selectorate goals—for example, whether they are purely seat-maximizing—our empirical predictions allow us to make partial inferences about these goals in real-world contexts.

A final institutional feature of the Swedish electoral system is a 1998 electoral reform that replaced closed lists with flexible lists, which endows voters with the option to cast a single preference vote for a candidate on any party's ballot. A candidate is elected ahead of her higher-ranked co-partisans if and only if she receives at least 5% of preference votes cast for her party. In practice—as with other countries that employ flexible list systems—candidates are almost never elected on the basis of preference votes, alone.

[9] Engström (2014)

[10] ibid.

5. Data and Measurement

Our data come from Sweden's administrative register, which covers every candidate on every electoral ballot in every election from 1982 to 2014. Variables in the dataset come from various government agencies such as the tax agency, the electoral agency, and the military defense ministry. We observe every candidate's rank on the ballot, and whether he or she was elected. These variables are perfectly measured and have no missing data. Although data is linked via the personal ID code of each individual in the dataset, this code is scrambled by Statistics Sweden and researchers work with the anonymized data.

In what follows, we describe how we create our measurement of ballot rank contestability, which then forms the basis for our sample restrictions. We also describe our four measurements of politician valence. To test our theoretical predictions, we need two key types of measurements: measurements of the contestability of different list ranks *and* measurements of politician valence. We first develop a novel methodology to compute rank contestability, and then discuss our construction of valence measures.

Computing contestability of ballot ranks. In closed and flexible party-list PR systems, a candidate's prospect of election increases with her list rank. We propose a data-driven and flexible approach to calculate the marginality of each rank within each party list. Our methodology reflects the idea that parties—and, possibly, voters—can use information about the electoral history of the party to predict its performance in the next election. For example, a party that has consistently won ten municipal seats is likely to win approximately ten seats. That party's fifth-ranked candidate occupies a relatively safe seat, while its twentieth-ranked candidate is almost sure to lose; neither seat is relatively likely to be contested. By contrast, the tenth and eleventh seats are more likely to hang in the balance. Using a simple regression framework, we predict the probability that each rank in each party list is converted into a seat.

We begin by constructing a dataset for the number of ranks won by each local party in each election. We then create a dummy variable, D^r_{pt}, for each ballot rank $r \in \{0, ..., 50\}$. This dummy takes the value 1 if the party p won this rank in the election t, and zero otherwise. We regress each of these dummies on all fifty dummies in the two most recent elections, $t-1$ and $t-2$. For the case of the eighth rank, for example, the equation becomes:

$$D^8_{pt} = \sum_{r=1}^{50} \delta^r D^r_{p,t-1} + X'_{p,t-1}\gamma + \sum_{r=1}^{50} \lambda^r D^r_{p,t-2} + X'_{p,t-2}\rho + \varepsilon_{p,t}. \tag{5}$$

Our setup incorporates additional predictors (X') that a party might use when calculating its expected number of seats. Our estimation accounts for the large variation in municipality size, and its consequence for the variability of seat shares: in a large municipality, a smaller percentage point vote shift is required for an additional seat than in a small municipality. We therefore interact council size with each rank-dummy and include these in X'. Additional and potential variables could include indicators for party size, for example approximated by seat shares. Equation (5) is estimated by OLS for clarity of interpretation, but recognizing that some predicted probabilities may fall outside of the unit range.

The predicted probability of winning eight seats is thus:

$$\hat{D}^8_{pt} = \sum_{r=1}^{50} \widehat{\delta}^r D^r_{p,t-1} + X'_{p,t-1}\widehat{\gamma} + \sum_{r=1}^{50} \widehat{\lambda}^r D^r_{p,t-2} + X'_{p,t-2}\widehat{\rho}. \tag{6}$$

Repeating this process for each of the fifty regressions gives us a new dataset of predicted probabilities for each list rank in each local party, and in each election. Note that we can choose which variables are included in equations (5) and (6)—for example, we can extend the dummies for the party's electoral history even further back in time. In the Swedish context, anecdotal evidence suggests that parties set the baseline for the expected number of seats at the number of seats in the previous election.[11]

To illustrate our approach, Figure 1 plots the averages of the predicted probabilities that a party wins each list rank, given that it won $k \in \{8, 10, 12, 14\}$ seats in the previous election: each line corresponds to each of the four hypothetical seat shares (i.e., each value of k) in the previous election. As expected, the smaller parties' prospects fall away at relatively higher ballot ranks. For each pre-election party size (k), the predicted probabilities show the expected decline over ballot ranks. We use the predicted win probabilities to categorize ballot ranks into six categories of electoral safety. The intention is, of course, to then compare the average valence of candidates across ranks. The point of departure for the categorization is the contested ranks, which we define as ranks with a win margin in the interval of 10% to 90%, illustrated in Figure 2. We further subdivide the contested ranks into "Advantaged Contested" ranks (70% to 90%) , "Highly Contested" ranks (30% to 70%), and "Disadvantaged Contested" ranks (10% to 30%). Our theoretical predictions about the quality of a candidate in each of the three contested ranks in our theoretical model can be interpreted as predictions about the *average* quality of candidates in each of the three categories of contested ranks.

We define the top name on the list as a separate category, the "*Capolista*".[12] We designate the interim ranks between the the capolista and the contested ranks to be "*Safe*". Finally, we

[11] Authors' interviews.

[12] The nomenclature is Italian for "top of the list".

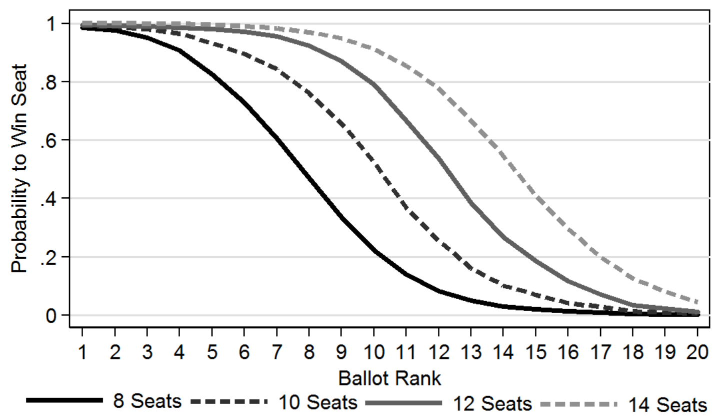

Figure 1 – **Average predicted win probabilities for ballot ranks**. The figure shows average predicted win probabilities (Y-axis) for each ballot rank (X-axis) and for political parties with either 8, 10, 12 or 14 seats. Probabilities are calculated by estimating equations (5) and (6) at the level of the party, election, and municipality.

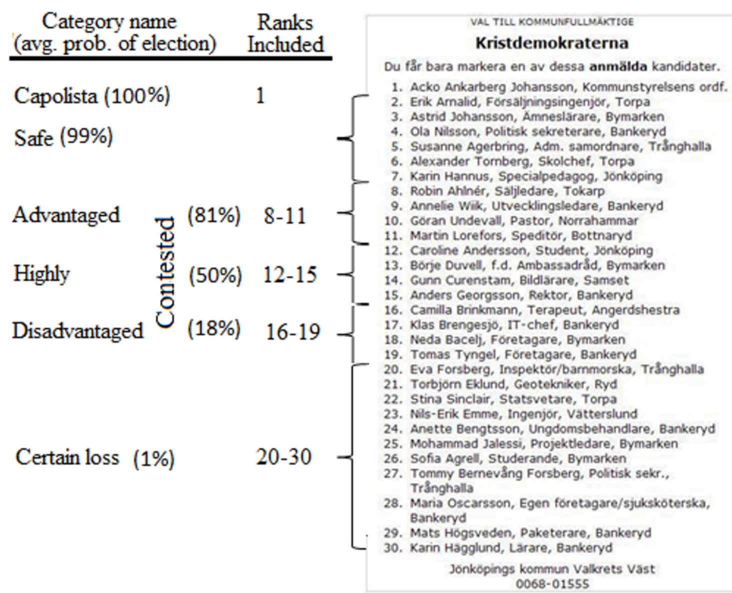

Figure 2 – **Ballot rank categorization of an example ballot**. The figure shows the six categories of ballot ranks for an example ballot of a party with 13 incumbent politicians. Category names are listed in the leftmost column together with the average predicted win probability in each category. Win probabilities are computed from OLS estimation of equations (5) and (6).

denote by "*Certain Loss*" the candidates with a predicted probability of election below ten percent. Figure 2 clarifies the six categories on a hypothetical ballot, in which the party won thirteen seats in the previous legislative cycle.

Our data contains 372,072 candidate-election observations and 15,819 municipal party-election observations. After restricting the data to local parties with at least one individual in each category, which we define as our estimation sample, we maintain 217,734 candidate-election observations and 6,561 municipal party-election observations. The municipal parties excluded are typically small: their median size is between 1 and 2 seats, with 14 percent of the excluded local parties winning no seats, and 90 percent winning 5 seats or less. The parties in our estimation sample have a median of 8 seats, and 75 percent have four seats or more. The estimation sample includes 64,085 of the 92,530 councilors elected during the period.

Measurements of candidate quality. Since there is no unified approach to quantifying the skills and experience—"valence"—that is at the heart of our theoretical framework, we use a broad approach that includes several non-controversial measurements. The first two come from data collected in Sweden's military enlistment—in particular, a mandatory draft that existed for men born between 1955-1979. The enlistment process included a test of cognitive abilities, as well as an interview to evaluate recruits' capacity for leadership, conducted by a trained psychologist. Both skills, cognitive and leadership, are normalized by the enlistment agency on a 1-9 stanine scale in the cohort, and are linked to our dataset via the personal ID code. For a detailed description of the enlistment variables and their strong correlations with private-sector careers, see Lindqvist and Vestman (2011) and Dal Bó *et al.* (2017).

A third, income-based valence measure was first suggested by Keane and Merlo (2010) and more thoroughly developed by Besley *et al.* (2017). It captures deviations in labor income among politicians with highly similar education, age and occupation. The idea is that an ideal measure of political competence should capture key governing abilities, but not be correlated with socioeconomic status. Politicians' earnings are therefore benchmarked against the totality of the Swedish working-age population in the same education-occupation-cohort and municipality cell. If politicians are paid a wage in their political roles, we use their earnings in the positions they held prior to becoming a full-time politician.[13]

We use education as a fourth and final valence variable. Recent related scholarly work has relied heavily on measures of candidate education, on the basis that it instills civic values and builds skills that improve political performance—see, for example, De Paola and Scoppa (2011); Besley and Reynal-Querol (2011); Galasso and Nannicini (2011); Ferraz and Finan (2009); Schwindt-Bayer (2011); Franceschet and Piscopo (2012) and Baltrunaite *et al.* (2014). We acknowledge, nonetheless, that education is a problematic measure of valence because of

[13] This measure is based on the idea that the private and political spheres are complements rather than substitutes. Besley *et al.* (2017) show that it is correlated with various measures of politician career success, voter support measured in preference votes, and that the proportion of municipal councilors who are defined as above-average in competence is correlated with better government outcomes.

its strong correlation with family background and social background.[14]

All four valence variables are standardized in pooled data for the 1991-2014 elections, and all local parties that meet our sample restrictions. The unit of measurement is in standard deviations from the variable mean. In an Appendix, we report raw means and standard deviations of the variables and preview their averages across our six categories of list ranks.

6. Empirical Results

Basic descriptives. Before testing our hypotheses, we first plot the relationship between list rank and candidate valence, according to the number of seats that a party won in the last election. In Figure 3 we plot the relationship between list rank and valence for parties that won an even number of seats between 2 and 8 seats in the last election.[15] For all party sizes and all 4 valence measures we see that candidate quality diminishes as we move further down the list ranks. For the two draft based measures there is is some variability, but this is likely due to the fact that we only have these measures for a smaller subset of the candidates.

Regression specification and baseline results. While the descriptive findings in Figure 3 would seem to cast doubt on the marginal rank hypothesis it does not provide a formal test. To formally test our two propositions we instead do the following. We first run a simple regression to quantify the average valence of politicians across the six categories of ballot ranks. Each of the four valence variables is regressed on dummy variables for the rank categories, excluding the "below marginal ranks" which serve as the reference.

$$Y_i = \alpha_{p,t} + \beta_1 Capo_i + \beta_2 Safe_i + \beta_3 Adv\ cont_i + \beta_4 Highly\ cont_i + \beta_5 Certain\ loss_i + \varepsilon_i. \quad (7)$$

Equation 7 also contains one dummy variable for each local party and election period. Our estimates hence capture average differences across rank categories within local parties, whilst isolating the comparison from variation across parties, municipalities, or even over time within the same local party.

The estimates on the dummy variables β_1 through β_5 capture the difference in average

[14] We observe a number of other candidate characteristics that could plausibly be related to quality, such as occupational background and political experience. We chose not use high-skilled occupations (e.g., law, business, or medicine) since it correlates more than education with family background and social class (Carnes 2013). Similarly, we exclude political experience: while capturing learned skills—bill drafting, bargaining and building coalitions—that are valuable in political organizations, it has been argued to measure "insider" status (Galasso and Nannicini, 2015). We also exclude local ties, being born and/or raised inside or outside the electoral district (e.g. Tavits, 2010, Shugart *et al.*, 2005), because of its lack of variation in our data.

[15] We present only even seat numbers to economize on space, but present odd numbers 1, 3, 5 and 7 in Figure S.2 in the Supplemental Appendix.

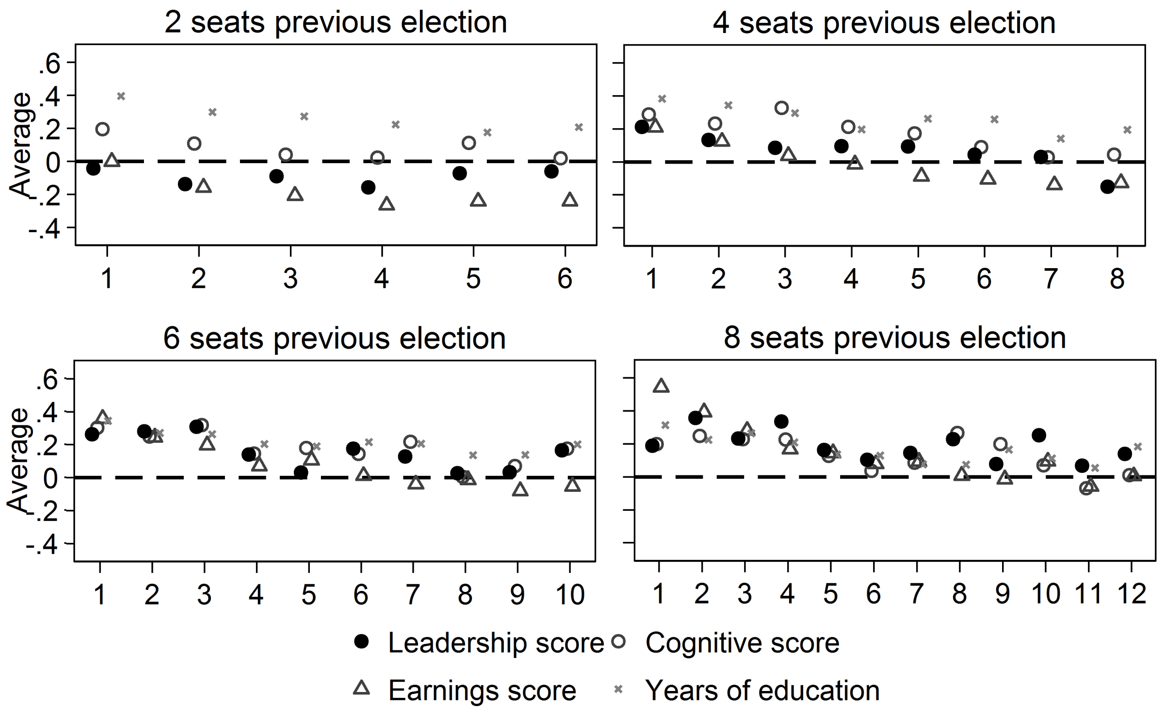

Figure 3 – **List rank and Valence**. The y-axis shows the average of each of the four valence measurements, transformed into Z-scores. Data includes all party groups between 1991 and 2014 that won the specified even number of seats in the previous election.

valence between politicians in each category and the reference category of *Disadvantaged Contested* ranks below the highly contested ranks. Most importantly, the estimates relate directly to the predictions of the model. Below, we explicitly state these predictions.

Predictions of the Marginal Rank Hypothesis. Unless a party is strongly disadvantaged, we should observe a non-monotonic relationship between candidate quality and ballot rank within a party's contested ranks. In particular:

(1) amongst parties that expect to win a relatively large seat share: $\max\{0, \hat{\beta}_4\} > \hat{\beta}_3$;
(2) amongst parties that expect to win a relatively small seat share: $\max\{\hat{\beta}_3, \hat{\beta}_4\} > 0$.

Predictions of the Top-Down Rank Order Hypothesis. Our estimates of equation (7) should satisfy:

$$\hat{\beta}_1 > \hat{\beta}_2 > \hat{\beta}_3 > \hat{\beta}_4 > 0 > \hat{\beta}_5. \tag{8}$$

Moreover, under Condition 2, the pattern should be the strongest in parties that have the highest probability of appointing the mayor.

Main Results. Figure 4 plots estimates the coefficients $\hat{\beta}_i$ in equation (7) for each of the four measures of candidate quality. The results clearly reject the marginal rank hypothesis, while

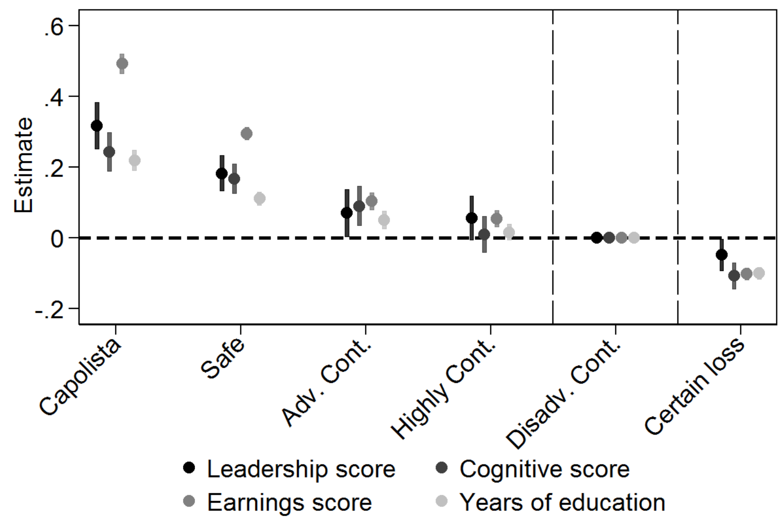

Figure 4 – **Average valence levels across ballot rank categories**. The figure shows estimated coefficients from Equation (7), where the average valence difference between each category of ballot ranks is estimated relative to the reference category of the disadvantaged competitive ranks. The outcome variables are the four valence measurements, transformed into Z-scores. Vertical lines show 95% confidence intervals. All regressions include fixed effects for every local party in every election period (3,950 dummies). The data is all candidates in local parties with at least one candidate in each of the six ballot rank categories. The number of observations in the regression analysis is: Leadership Score: $N = 35,662$; Cognitive Score: $N = 48,088$, Earnings Score: $N = 209,995$, Years of Education: $N = 185,382$.

providing strong support for the top-down rank order hypothesis. Across all party lists, i.e., in both closed and flexible list contexts, averaging across all parties and municipalities in our estimation sample, and across all four measures of candidate quality, average valence consistently increases with electoral security. In fact, there is a not a single instance in which average candidate quality increases as we move down the ballot rank categories.

We provide the details and results of a number of sensitivity tests in Supplemental Appendix D, including:

- splitting our data according to closed (pre-1998) and flexible (post-1998) lists (Figure S.3);
- changing the information that predicts each rank's election prospects, i.e., changing the variables that are included in X' in equation (5), (Figure S.9);
- using alternative thresholds to define our list rank categories (Figure S.7);
- using a candidate's list rank relative to the number of seats won in the previous election, instead of probability-based list rank categories (Figure S.8); and,

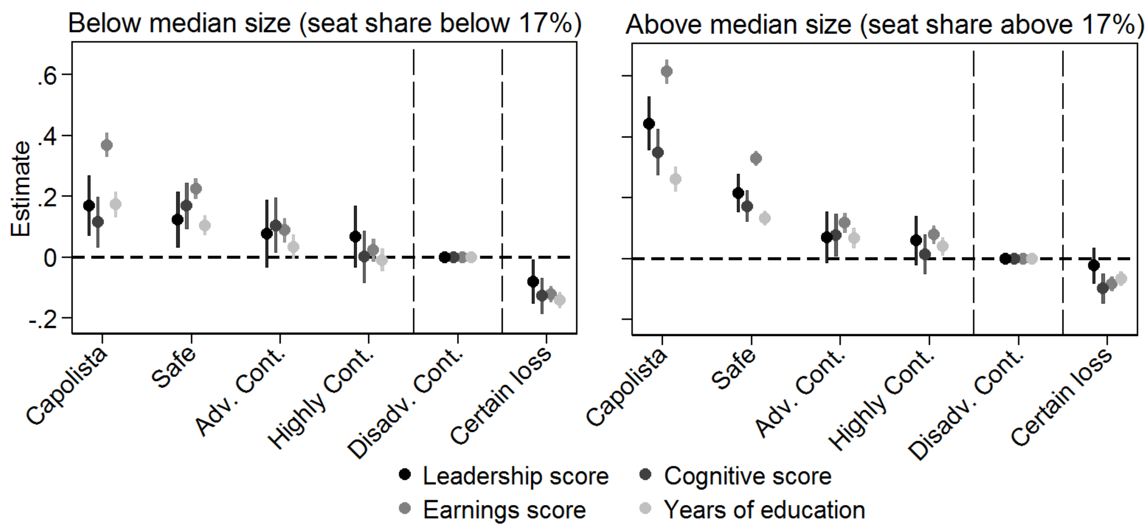

Figure 5 – **Average valence levels across ballot rank categories in *small* and *large* seat share political parties**. The figure replicates the analysis in Figure 4 in two sub-samples of data, split by median party size.

- recognizing that Swedish parties differ in their nomination procedures—for example, left-wing parties are more likely to employ primaries—we split our sample according to party ideology (Figure S.11).

Irrespective of sensitivity test that we employ, we clearly reject the marginal rank hypothesis, and obtain strong support for the top down rank hypothesis.

Testing ancillary empirical predictions. Proposition 1 predicts that the strongest prospect of observing a *non-monotonic* relationship between candidate quality and list rank within a party's contested ranks will be found amongst parties that are expected to win a large share of seats. In stark contrast, Proposition 2 states that the strongest prospect of observing a *monotonic* relationship between candidate quality and list rank within a party's contested ranks will be found amongst parties that are expected to win a large share of seats.

To examine these competing predictions, directly, Figure 5 divides party lists into two categories: those which, in their respective districts, enjoy a seat share above and below 17% at the time of the election—the median seat share in our sample. Across each of the rank categories and valence measures, the pattern is the same: the average valence amongst the advantaged contested ranks strictly exceeds the average valence in *each* of the remaining contested ranks. In fact, the correlation between rank and valence is *even stronger* in relatively larger seat share parties than it is in parties with a relatively smaller seat share—decisively rejecting the predictions of the marginal rank hypothesis, but consistent with the top-down rank order hypothesis.[16]

7. Alternative Explanations

While other theories could also explain why parties place better candidates in higher list ranks, these alternative theories must also explain why this relationship is more pronounced in electorally strong, versus electorally weak political parties. We discuss some of these alternative theories, briefly, relegating additional theoretical and empirical analysis to a Supplemental Appendix.

Different Candidate Pools Across Advantaged versus Disadvantaged Parties. Electorally advantaged parties might have access to a better supply of very high quality candidates. In that case, Figure 5 could simply reflect differences in the average quality of candidates at the top end of a party's ballot. In Supplemental Appendix D, we reproduce Figure 5 but plot our raw mean quality measures, instead of our estimates of equation (7). We uncover no evidence that the average quality of candidates in higher rank categories is greater in advantaged parties. We also develop and implement an alternative statistical test that is robust to differences in the quality of different parties' candidate pools. All of our findings extend.

Competition for the Top Spot. Perhaps electorally advantaged parties generate more intense competition for the top ranks amongst the parties' candidates, since these parties offer a relatively stronger prospect of wielding valuable executive power. If, in addition, a candidate's effort and quality are complements in the contest for better rank assignments, we would therefore expect higher quality candidates to appear at the top end of the ballot more often in electorally strong parties.[17]

However, while relatively strong parties can offer their highest-ranked politician the most attractive post-election appointment, a candidate's incentives to compete for this spot are offset by the fact that even lesser-ranked politicians in strong parties are also likely to enjoy privileged positions in the assembly. For example, the party that appoints the mayor is also likely to appoint the vice-mayor, and committee chairs. By contrast, electorally weaker parties that are likely to hold minority status may have at most one high-status position to offer to its elected politicians. anyone outside the very highest spot may face a significant risk of

[16] In Supplemental Appendix D, we consider an alternative split of our sample, by dividing districts into *above* and *below* median competition, where competition is defined as the average vote share difference between the two party blocs over the previous three elections. We report estimates of equation (7) for this alternative sample split, and our conclusions are unchanged.

[17] Recall, however, that we observe a stronger relationship between ballot rank and candidate quality amongst advantaged parties at *all* ballot levels.

non-election, or be consigned to rank-and-file status in the minority.

More generally: competition for a top spot in a small party is more likely to be *winner-take-all*; strong parties, by contrast, are able to offer a more progressive schedule of rewards even outside of the very highest ranks. This logic would imply that the most intense competition for top ranks should be found within electorally *weak* parties—generating a prediction that contradicts both our theoretical predictions in Proposition 2 and our empirical results.

Better candidates may also be better able to threaten to revert to another party. This explanation could certainly play an important role in political contexts where party competition is primarily distributive, e.g., Brazil. But it is less likely to apply in contexts where parties are founded and differentiated on ideological grounds, such as Sweden. In these contexts, party-switching is extremely rare.

Voter attention and candidate visibility. Our theoretical results incorporate assumptions about both parties' goals and a perspective on how voters formulate their induced preferences over party lists, i.e., according to Lemma 2. Our framework does not require that *all* voters pay attention to individual candidates, but rather that *some* voters are responsive.

To what extent is this view reasonable? Perhaps all voters are poorly informed about party lists, and focus deterministically only on candidates that are placed in safe ranks at the top of the ballot. In our framework, this is equivalent to replacing the equilibrium attention weights $\tau(k)$ for each rank k in Lemma 2 with exogenous weights satisfying $\tau(1) > \tau(2) \geq ...\tau(5)$, with, $\tau(k) = 0$ for some $2 \leq k \leq 4$.

Sweden's introduction of flexible lists in 1998 offers an opportunity to examine this possibility. The reform allowed each voter to cast a single voluntary preference vote. In the ballot booth, a pen is available to cast this optional preference vote by checking a box next to the preferred politician's name. If no box is checked, the voter does not influence the internal ranking of the party. The proportion of preference voters is roughly equal across categories of sex, age, education and occupational groups. Our data includes the tally of preference votes for every politician in every rank in each of the five local elections that were held between 1998 and 2014, i.e., after the reform.

In the left panel of Figure 6 we plot estimates of equation (7), with the proportion of preference votes as the outcome variable. The right panel highlights pairwise correlations between preference votes and valence among politicians in each of our six ballot rank categories. The regression that captures the pairwise correlations includes fixed effects for the interaction between local party, election, and municipality. This restricts our variation to voter behavior within a given list in a given election, meaning also that we cannot get an estimate for top-

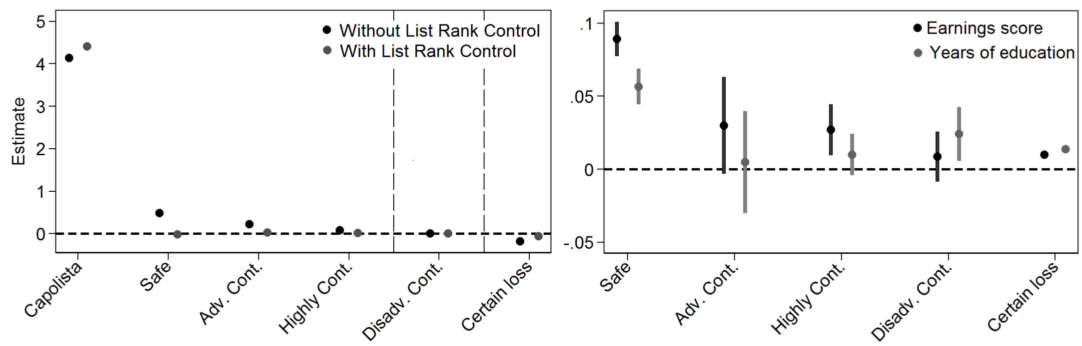

Figure 6 – **Preference votes relative to disadvantaged contested seats (bottom left) and pairwise correlations between valence and preference votes in each ballot rank category (bottom right)**. Each figure uses data for parties with at least one politician in each ballot rank category and data for all elections between 1998 and 2014. The left-hand side figure reports estimated coefficients with 95% confidence intervals from equation (7) where the "disadvantaged contested ranks" are the reference category, and the share of preference votes is the outcome variable. The regression includes fixed effects for every local party in every election period, and is estimated both with and without fixed effects for ballot rank. The right-hand side figure shows estimates from an OLS regression for the relationship between valence and preference votes; and using the two valence measures available for the full population (earnings score and education length). A separate regression is estimated for each category of ballot ranks and each valence measure, all including fixed effects for every local party in every election period.

ranked politicians (as there is only a single such candidate in a list). Both the fixed effects and data limitation to a subset of male cohorts generate substantial noise in our military enlistment variables, and we omit these.

The vast majority of preference votes are indeed awarded to the top portion of party lists. Consistent with the extension of our framework to flexible list settings—detailed in the Supplemental Appendix C—the top-ranked politician alone receives about a third of the total votes in the average party list. More generally, the top three politicians, together, account for approximately one half of all votes. On the one hand, this is consistent with the idea that the highest ranks are focal for voters.

On the other hand, the right panel highlights that the correlation between candidate quality and preference *remains substantial* as one moves out of the safe seats, i.e., the highest ranks. Preference votes are still responsive to quality within the contested ranks. So, while voters generally cast their ballots for higher-ranked candidates *and* their preference votes are also more sensitive to candidate quality within higher ranks, their votes are also responsive to variation in candidate quality outside of these ranks. This casts doubt on the empirical validity of a simple heuristic in which voters solely focus on the very highest ranks. And, even if voters were guided by this simple heuristic, this would not account for the differentially strong relationship between between rank and quality in strong versus weak parties.

8. Conclusion

We analyze party list assignments both in theory and in real-world contexts. We develop a theoretical framework to examine party list assignments in multi-member contexts, under alternative specifications of how party goals are facilitated by the conjunction of legislative seats and the skills of individual politicians. We test two predictions generated by the framework, leveraging a dataset containing the universe of Swedish politicians on every ballot in 290 municipal elections over seven election periods.

A critical premise underpinning the existing literature on candidate selection is that voters care about candidate quality more than parties. For instance, party elites may prefer to select candidates with a view to their ideological congruence (Serra, 2011), or their loyalty (Galasso and Nannicini, 2015), amongst many plausible non-quality criteria. Our results suggest a more optimistic view of the role that parties can play in fostering effective political selection. We find that it is the parties that face the weakest electoral constraints that are most likely to place better candidates in better ranks. This finding cannot be rationalized by an account in which the demand for quality arises solely or even predominantly from voters.

Our framework is intended to facilitate the development of theoretical hypotheses that are applicable to many different list PR contexts—both within and across countries. Nonetheless, we recognize that this requires us to abstract from subtle and important variation in the way these electoral rules operate in real-world contexts, which is explored in existing and ongoing scholarly efforts that focus on particular country-contexts, such as Cheibub and Sin (2018). We hope that our attempt to develop a broad theoretical framework will play a complementary role in shedding further theoretical and empirical light on candidate selection in this important class of electoral rules.

References

André, A., Depauw, S., Shugart, M. S. and Chytilek, R. (2015). Party nomination strategies in flexible-list systems: Do preference votes matter? *Party Politics*, **23**, 589–600.

Baltrunaite, A., Bello, P., Casarico, A. and Profeta, P. (2014). Gender quotas and the quality of politicians. *Journal of Public Economics*, **118**, 62–74.

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

— and REYNAL-QUEROL, M. (2011). Do democracies select more educated leaders? *American Political Science Review*, **105** (3), 552–566.

CHEIBUB, J. and SIN, G. (2018). Intraparty competition and open-list pr elections: What can we learn from an extremely permissive system?, http://tiny.cc/vtjw9y.

CRUTZEN, B., FLAMAND, S. and SAHUGUET, N. (2017). Prize allocation and incentives in team contests, http://sabineflamand.com/research/Teams.pdf.

CRUTZEN, B. S. and SAHUGUET, N. (2017). Electoral incentives: the interaction between candidate selection and electoral rules, http://tiny.cc/izjw9y.

CRUZ, C., KEEFER, P. and SCARTASCINI, C. (2016). The database of political institutions 2015 (dpi2015). *Inter-American Development Bank. Numbers for Development.* *https://publications.iadb.org/en/database-political-institutions-2017-dpi2017*.

DAL BÓ, E. and FINAN, F. (2018). Progress and perspectives in the study of political selection. *Annual Review of Economics*, **10**, 541–575.

—, FINAN, F., FOLKE, O., PERSSON, T. and RICKNE, J. (2017). Who becomes a politician? *The Quarterly Journal of Economics*, **132** (4), 1877–1914.

DE PAOLA, M. and SCOPPA, V. (2011). Political competition and politician quality: evidence from italian municipalities. *Public Choice*, **148** (3), 547–559.

DOWNS, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, **65** (2), 135–150.

DUVERGER, M. (1959). *Political parties: Their organization and activity in the modern state*. Methuen.

ENGSTRÖM, M. (2014). *Vägen till valsedeln, Riksdagspartiernas nomineringsprocesser och dess betydelse*. Ph.D. thesis, Lund University.

ENGSTROM, R. L. (1991). *Presidential Studies Quarterly*, **21** (1), 167–168.

ESTEVE-VOLART, B. and BAGUES, M. (2012). Are women pawns in the political game? evidence from elections to the spanish senate. *Journal of Public Economics*, **96** (3-4), 387–399.

FEARON, J. D. (1999). Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance. *Democracy, accountability, and representation*, **55**, 61.

FERRAZ, C. and FINAN, F. (2009). Motivating politicians: The impacts of monetary incentives on quality and performance, https://www.nber.org/papers/w14906.

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

FRANCESCHET, S. and PISCOPO, J. M. (2012). Gender and political backgrounds in argentina. In *The Impact of Gender Quotas*, Oxford University Press New York, pp. 43–56.

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

— and — (2015). So closed: Political selection in proportional systems. *European Journal of Political Economy*, **40**, 260–273.

HIRANO, S. and SNYDER, J. (2014). Primary elections and the quality of elected officials. *Quarterly Journal of Political Science*, **9** (4), 473–500.

HOBOLT, S. B. and HØYLAND, B. (2011). Selection and sanctioning in european parliamentary elections. *British Journal of Political Science*, **41** (3), 477–498.

KEANE, M. P. and MERLO, A. (2010). Money, political ambition, and the career decisions of politicians. *American Economic Journal: Microeconomics*, **2** (3), 186–215.

KEY, V. O. (1964). *Politics, Parties, and Pressure Groups*. Thomas Y. Crowell Company.

LINDQVIST, E. and VESTMAN, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, **3** (1), 101–128.

LUNDQVIST, H. (2013). Is it worth it? on the returns to holding political office, https://www.hhs.se/contentassets/290e656dbfbc407fa9d5ef9bc6700642/hlundqvist_politicalreturns.pdf.

MATAKOS, K., SAVOLAINEN, R., TROUMPOUNIS, O., TUKIAINEN, J. and XEFTERIS, D. (2018). Electoral institutions and intraparty cohesion. VATT Institute for Economic Research, https://www.doria.fi/handle/10024/159572.

MERILÄINEN, J. and TUKIAINEN, J. (2018). Rank effects in political promotions. *Public Choice*, pp. 1–23.

MONTIN, S. (2015). Municipalities, regions, and county councils. *The Oxford Handbook of Swedish Politics*, p. 367.

NEMOTO, K. and SHUGART, M. S. (2013). Localism and coordination under three different electoral systems: The national district of the japanese house of councillors. *Electoral Studies*, **32** (1), 1–12.

PATTY, J., SCHIBBER, C., PENN, M. and CRISP, B. (Forthcoming). Valence, elections, and legislative institutions. *American Journal of Political Science*.

SCHWINDT-BAYER, L. A. (2011). Women who win: Social backgrounds, paths to power, and political ambition in latin american legislatures. *Politics & Gender*, **7** (1), 1–33.

SERRA, G. (2011). Why primaries? the party's tradeoff between policy and valence. *Journal of Theoretical Politics*, **23** (1), 21–51.

SHUGART, M. S., VALDINI, M. E. and SUOMINEN, K. (2005). Looking for locals: voter information demands and personal vote-earning attributes of legislators under proportional representation. *American Journal of Political Science*, **49** (2), 437–449.

SLOUGH, T., YORK, E. and TING, M. (2018). A dynamic model of primaries. *Journal of Politics*, forthcoming.

SNYDER, J. M., JR and TING, M. M. (2011). Electoral selection with parties and primaries. *American Journal of Political Science*, **55** (4), 782–796.

TAVITS, M. (2010). Effect of local ties on electoral success and parliamentary behaviour: The case of estonia. *Party Politics*, **16** (2), 215–235.

WEINGAST, B. R. and MARSHALL, W. J. (1988). The industrial organization of congress; or, why legislatures, like firms, are not organized as markets. *Journal of Political Economy*, **96** (1), 132–163.

Supplementary Appendix to *Party Nomination Strategies in List Proportional Representation Systems* (For Online Publication)

Table of Contents

Appendix A: Proofs of Propositions

For ease of exposition, we include the statements of each result.

Lemma 2 *For any party lists l_A and l_B, there exists a pair $V_A(l_A, l_B)$ and $V_B(l_A, l_B)$ such that an instrumental voter's value from party B is:*

$$V_A(l_A, l_B) = \sum_{k=2}^{4} \tau(k)\ \tilde{H}(k, l_A, l_B),$$

while her value from party A is:

$$V_B(l_A, l_B) = \sum_{k=2}^{4} \tau(k)\ \tilde{H}(k-1, l_A, l_B).$$

Moreover, when θ is small enough, for each pair of lists (l_A, l_B), the difference $\Delta(l_A, l_B) = V_A(l_A, l_B) - V_B(l_A, l_B)$ is uniquely determined.

Proof of Lemma 2. For a given voter, the probability of being pivotal for the assignment of A's k^{th} seat ($k \in \{2, 3, 4\}$), conditional on being pivotal, and given other voters' computed $\Delta \equiv V_A - V_B$ is given by

$$\frac{\Pr\left(\pi_A = \pi(k)\right)}{\sum_{r=2}^{4} \Pr\left(\pi_A = \pi(r)\right)} = \frac{f\left(\phi^{-1}\pi(k) - \phi^{-1}\frac{1}{2} - \Delta\right)}{\sum_{r=2}^{4} f\left(\phi^{-1}\pi(r) - \phi^{-1}\frac{1}{2} - \Delta\right)} = \frac{f\left(\tilde{\pi}(k) - \Delta\right)}{\sum_{r=2}^{4} f\left(\tilde{\pi}(r) - \Delta\right)} \equiv \tau_k(\Delta)$$

where $\tilde{\pi}(k) = \phi^{-1}\left(\pi(k) - \frac{1}{2}\right)$ is such that, by symmetry $\sum_{r=2}^{4} \tilde{\pi}(k) = 0$. We can then rewrite the value of voting for A for a voter given other voters' computed $V_A - V_B$ as follows:

$$V_A(\Delta) = \sum_{k=2}^{4} \tau_k(\Delta) \tilde{H}(k; l_A, l_B) \tag{S.1}$$

$$V_B(\Delta) = \sum_{k=2}^{4} \tau_k(\Delta) \tilde{H}(k-1; l_A, l_B) \tag{S.2}$$

First, we show that there exists $\kappa > 0$ such that when $|\tilde{f}'(\cdot)| \leq \kappa$, the mapping $\mathcal{V} : [\underline{V} - \overline{V}, \overline{V} - \underline{V}] \to [\underline{V} - \overline{V}, \overline{V} - \underline{V}]$ defined as

$$\mathcal{V}(\Delta) = \sum_{k=2}^{4} \tau_k(\Delta) \left[\tilde{H}(k; l_A, l_B) - \tilde{H}(k-1; l_A, l_B)\right]$$

has a unique fixed point. To see this, first notice that $\tilde{H}(k;l_A,l_B) - \tilde{H}(k-1;l_A,l_B) < \overline{V} - \underline{V}$ (since the two values differ only by one elected councilor) and $\mathcal{V}$ is a convex combination of three such values, which implies that $\mathcal{V}(\underline{V} - \overline{V}) > \underline{V} - \overline{V}$ and $\mathcal{V}(\overline{V} - \underline{V}) < \overline{V} - \underline{V}$. Second, notice that, by construction, $\sum_{k=2}^{4} \tau_k(\Delta) = 1$, which implies that (suppressing the dependence of $\tilde{H}(k;l_A,l_B)$ on (l_A,l_B) for clarity)

$$\mathcal{V}(\Delta) = \tau_2(\Delta)\big[2\tilde{H}(2) - \tilde{H}(1) - \tilde{H}(3)\big] + \tau_4(\Delta)\big[\tilde{H}(4) + \tilde{H}(2) - 2\tilde{H}(3)\big] + \tilde{H}(3) - \tilde{H}(2)$$

Hence, we can write

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\Delta}\mathcal{V}(\Delta) &= [2\tilde{H}(2) - \tilde{H}(1) - \tilde{H}(3)]\frac{\mathrm{d}}{\mathrm{d}\Delta}\tau_2(\Delta) + [\tilde{H}(4) + \tilde{H}(2) - 2\tilde{H}(3)]\frac{\mathrm{d}}{\mathrm{d}\Delta}\tau_4(\Delta) \\ &< (\overline{V} - \underline{V})\left(\left|\frac{\mathrm{d}}{\mathrm{d}\Delta}\tau_2(\Delta)\right| + \left|\frac{\mathrm{d}}{\mathrm{d}\Delta}\tau_4(\Delta)\right|\right) < \frac{\kappa\theta(\overline{V} - \underline{V})}{\sum_{r=2}^{4} f\left(\tilde{\pi}(r) - \Delta\right)}[2 + \tau_2(\Delta) + \tau_4(\Delta)]\end{aligned}$$

since

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}\Delta}\tau_k(\Delta)\right| &= \left|\frac{f\left(\tilde{\pi}(k) - \Delta\right)\sum_{r\neq k} f'\left(\tilde{\pi}(r) - \Delta\right) - f'\left(\tilde{\pi}(k) - \Delta\right)\sum_{r\neq k} f\left(\tilde{\pi}(r) - \Delta\right)}{\left(\sum_{r=2}^{4} f\left(\tilde{\pi}(r) - \Delta\right)\right)^2}\right| \\ &< \frac{\kappa\theta f\left(\tilde{\pi}(k) - \Delta\right) + \kappa\theta\sum_{r=2}^{4} f\left(\tilde{\pi}(r) - \Delta\right)}{\left(\sum_{r=2}^{4} f\left(\tilde{\pi}(r) - \Delta\right)\right)^2} = \frac{\kappa\theta}{\sum_{r=2}^{4} f\left(\tilde{\pi}(r) - \Delta\right)}[1 + \tau_k(\Delta)]\end{aligned}$$

and $f'(x) = \theta\tilde{f}'(x)$ and $\kappa \equiv \max \tilde{f}'(x)$, $x \in \left[-\frac{1}{2\psi}, \frac{1}{2\psi}\right]$. As a consequence, whenever θ is small enough, $\frac{\mathrm{d}}{\mathrm{d}\Delta}\mathcal{V}(\Delta) < 1$ and $\mathcal{V}(\Delta)$ admits a unique fixed point. □

Lemma ??. If party A is sufficiently advantaged, i.e., $\overline{\xi}$ is large enough, then $\tau(4) > \tau(3) > \tau(2)$ for all $(l_A,l_B) \in L(Q)^2$.

Proof. Let $\overline{\xi} \geq \overline{V} - \underline{V} + \tilde{\pi}(4)$. In this case, we have that $\forall\ \Delta \in (\underline{V} - \overline{V}, \overline{V} - \underline{V})$,

$$\tilde{\pi}(2) - \Delta < \tilde{\pi}(3) - \Delta < \tilde{\pi}(4) - \Delta < \overline{\xi}$$

which implies

$$f(\tilde{\pi}(2) - \Delta) < f(\tilde{\pi}(3) - \Delta) < f(\tilde{\pi}(4) - \Delta)$$

This completes the proof. □

Proposition 1. (Marginal Rank Hypothesis) Under closed-list PR, when parties are purely seat-motivated, for θ not too large:

1. both parties place their best candidates in their competitive ranks;
2. within the competitive ranks, A's list does not put its strongest candidate at the top and B's list does not place its highest quality candidate at the bottom:

$$l_A(2) < \max\{l_A(3), l_A(4)\}\,; \quad l_B(4) < \max\{l_B(2), l_B(3)\}.$$

3. When A's advantage, i.e., $\bar{\xi}$, is large enough, then for any pair of contested ranks $k' > k$,

$$l_A(k') > l_A(k)\,; \quad l_B(k') < l_B(k).$$

Note that our statement of Proposition 1 in the main text makes reference only to our result about the ordering of candidates within the contested ranks, i.e., points 2 and 3. In Supplemental Appendix C, we extend these results directly to flexible lists. We also prove additional results about the quality of candidates outside of the contested ranks, i.e., an analogue of point 1.

Proof. From Lemma 2, the mapping $\mathcal{V}$ admits a unique fixed point and thus there exists a unique $\Delta(l)$ for each pure strategy profile (l_A, l_B). Let λ_J denote a mixed strategy for party J (a probability distribution over the elements of $L(Q)$) and $\mathcal{L}$ denote the set of individually rational mixed strategy profiles. An equilibrium is a profile $(\lambda_A^*, \lambda_B^*)$ that solves

$$\min_{\lambda_B} \max_{\lambda_A} \sum_{y \in L(Q)} \sum_{z \in L(Q)} \Delta(y, z) \lambda_A(y) \lambda_B(z)$$

Since the set $L(Q)$ is finite, by the Minimax Theorem, an equilibrium exists, and it produces a unique Δ^*. This immediately implies that there is a unique vector of attentions weights $\{\tau^*(2), \tau^*(3), \tau^*(4)\}$.

Part 1. We can rewrite the mapping $\mathcal{V}$ implicitly defining Δ as follows:

$$\mathcal{V}(\Delta) = \frac{\tilde{H}(4; l_A, l_B) - \tilde{H}(1; l_A, l_B)}{3} + \sum_{k=2}^{4} \varepsilon_k(\Delta) \big[\tilde{H}(k; l_A, l_B) - \tilde{H}(k-1; l_A, l_B)\big]$$

where $\varepsilon_k(\Delta) = \tau_k(\Delta) - \frac{1}{3}$. Since $\varepsilon_k(\Delta) \underset{\theta \to 0}{\longrightarrow} 0$, there exists θ small enough for which

$$\tilde{H}(4; l_A, l_B) - \tilde{H}(1; l_A, l_B) > (<)0 \Rightarrow \Delta(l_A, l_B) > (<)0$$

Notice that, for any pair (l_A, l_B):

$$\tilde{H}(4; l_A, l_B) - \tilde{H}(1; l_A, l_B) = H\left(\sum_{i=1}^{4} l_A(i) + l_B(1), \max_{i\leq 4} l_A(i)\right) - H\left(l_A(1) + \sum_{i=1}^{4} l_B(i), \max_{i\leq 4} l_B(i)\right). \tag{S.3}$$

We next claim that each party must put its worst candidate, i.e., with quality 1, in an uncontested rank, when θ is small enough. We prove the argument for party A. Suppose not, i.e., that a list l_A is chosen with positive probability that satisfies $1 \in \{l_A(2), l_A(3), l_A(4)\}$; let $i \in \{2, 3, 4\}$ denote the rank such that $l_A(i) = 1$. Consider an alternative list l'_A that sets $l'_A(5) = 1$ and $l'_A(i) = l_A(5)$ and otherwise replicates l_A. For any realized l_B, we have that $\tilde{H}(1; l_A, l_B) = \tilde{H}(1; l'_A, l_B)$, and also $\max_{j\leq 4} l'_A(j) = 5$, and thus:

$$\begin{aligned}
&\tilde{H}(4; l'_A, l_B) - \tilde{H}(1; l'_A, l_B) - [\tilde{H}(4; l_A, l_B) - \tilde{H}(1; l_A, l_B)] \\
=&H\left(\sum_{j=1}^{4} l_A(j) + \underbrace{l_A(5) - l_A(i)}_{>0} + l_B(1), 5\right) - H\left(\sum_{i=1}^{4} l_A(i) + l_B(1), \max_{i\leq 4} l_A(i)\right) > 0,
\end{aligned} \tag{S.4}$$

where $\max_{i\leq 4} l_A(i) \leq 5$. Since (S.4) holds for *any* list l_B, we conclude that l_A is strictly dominated, for θ small.

We next claim that each party must put its second-worst candidate, i.e., with quality 2, in an uncontested rank, when θ is small enough. We prove the argument for party A. Suppose not, i.e., that a list l_A is chosen with positive probability satisfying $2 \in \{l_A(2), l_A(3), l_A(4)\}$; let $i \in \{2, 3, 4\}$ denote the rank such that $l_A(i) = 2$. There are two cases, depending on whether $l_A(1) = 1$, or instead $l_A(5) = 1$.

Case 1: $l_A(1) = 1$. Consider an alternative list l'_A that sets $l'_A(5) = 2$ and $l'_A(i) = l_A(5)$ and otherwise replicates l_A. For any realized l_B, we have that $\tilde{H}(1; l_A, l_B) = \tilde{H}(1; l'_A, l_B)$, and also $\max_{j\leq 4} l'_A(j) = 5$, and thus the difference $\tilde{H}(4; l'_A, l_B) - \tilde{H}(1; l'_A, l_B) - [\tilde{H}(4; l_A, l_B) - \tilde{H}(1; l_A, l_B)]$ is again given by expression (S.4) for *any* list l_B. We conclude that l_A is strictly dominated, for θ small.

Case 2: $l_A(1) \neq 1$. Then, our previous claim implies that $l_A(5) = 1$, and thus $l_A(1) > 2$. Consider an alternative list l'_A that sets $l'_A(1) = 2$ and $l'_A(i) = l_A(1)$. For any realized l_B, we have that $\tilde{H}(4; l'_A, l_B) - \tilde{H}(4; l_A, l_B) = 0$, and thus

$$\begin{aligned}
&\tilde{H}(4; l'_A, l_B) - \tilde{H}(1; l'_A, l_B) - [\tilde{H}(4; l_A, l_B) - \tilde{H}(1; l_A, l_B)] \\
=& \tilde{H}(1; l_A, l_B) - \tilde{H}(1; l'_A, l_B)
\end{aligned}$$

$$= \tilde{H}\left(l_A(1) + \sum_{i=1}^{4} l_B(i), \max_{j\leq 4} l_B(j)\right) - \tilde{H}\left(2 + \sum_{i=1}^{4} l_B(i), \max_{j\leq 4} l_B(j)\right) > 0, \tag{S.5}$$

since $l_A(1) > 2$. Since this inequality holds for *any* list l_B, we therefore conclude that l_A is strictly dominated, for θ small. This proves that, for $\theta > 0$, a list l_J is not strictly dominated only if $l_J(1) \in \{1, 2\}$ and $l_J(5) \in \{1, 2\}$.

We conclude that, for any pair (l_A, l_B) that arises with positive probability, if θ is small enough:

$$\begin{aligned}
&\tilde{H}(4; l_A, l_B) - \tilde{H}(1; l_A, l_B) \\
&\quad = H(l_A(1) + l_B(1) + \underbrace{3+4+5}_{=\sum_{r=2}^{4} l_A(r)}; 5) - H(l_A(1) + l_B(1) + \underbrace{3+4+5}_{=\sum_{r=2}^{4} l_B(r)}; 5) = 0,
\end{aligned} \tag{S.6}$$

i.e., Δ^* is close enough to zero, such that for every $\delta > 0$ we can find a θ small enough so that for *any* list (l_A, l_B) that is played with positive probability:

$$\tau_k(\Delta(l_A, l_B)) - \tau_k(0) < \delta.$$

This, in turn, implies that for θ small enough, $\tau_2(\Delta(l_A, l_B)) < \tau_3(\Delta(l_A, l_B))$ and $\tau_2(\Delta(l_A, l_B)) < \tau_4(\Delta(l_A, l_B))$.

We now show that a list l_A satisfying $l_A(2) = 5$ is a strictly dominated action for party A. Suppose that in equilibrium $l_A(2) = 5$ with positive probability and consider an alternative list l'_A that sets $l'_A(2) = l_A(3)$ and $l'_A(3) = l_A(2)$ and otherwise satisfies $l_A(k) = l'_A(k)$. It is immediate to see that for any list l_B and rank $k \in \{3, 4\}$, $\tilde{H}(k; l_A, l_B) = \tilde{H}(k; l'_A, l_B)$. First, we observe that:

$$\begin{aligned}
&\tilde{H}(2; l'_A, l_B) - \tilde{H}(2; l_A, l_B) \\
=&H\left(l_A(1) + l_A(2) + \underbrace{l_A(3) - l_A(2)}_{<0} + \sum_{j=1}^{3} l_B(j), 5\right) - H\left(l_A(1) + l_A(2) + \sum_{j=1}^{3} l_B(j), 5\right) < 0.
\end{aligned}$$

Second, we show that $\Delta(l_A, l_B)$ is strictly decreasing in $\tilde{H}(2; l'_A, l_B)$, which implies that $\Delta(l'_A, l_B) > \Delta(l_A, l_B)$, completing the argument. Notice that, using the fact that in any equilibrium $\tilde{H}(4; l_A, l_B) = \tilde{H}(1; l_A, l_B)$, we can rewrite $\mathcal{V}(\Delta)$ as

$$\begin{aligned}
\mathcal{V}(\Delta) =&\tilde{H}(4; l_A, l_B)[\tau_4(\Delta) - \tau_2(\Delta)] \\
&+\tilde{H}(3; l_A, l_B)[\tau_3(\Delta) - \tau_4(\Delta)]
\end{aligned}$$

$$+\tilde{H}(2;l_A,l_B)[\tau_2(\Delta)-\tau_3(\Delta)]. \tag{S.7}$$

Hence, by the implicit function theorem,

$$\frac{d\Delta(l_A,l_B)}{d\tilde{H}(2;l_A,l_B)}=\frac{\tau_2(\Delta)-\tau_3(\Delta)}{1-\frac{\partial\nu}{\partial\Delta}}<0$$

which follows form $\tau_2(\Delta)-\tau_3(\Delta)<0$, which we established earlier, and the the fact that, by Lemma 2, $\frac{\partial\nu}{\partial\Delta}<1$. The argument for why $l_B(4)<5$ follows a similar logic.

Part 2. For each θ satisfying the assumptions of parts 1-3, we may find $\overline{\xi}$ large enough so that for every pair of individually rational lists (l_A,l_B), $\tilde{\pi}(4)+\Delta(l_A,l_B)<\overline{\xi}$. This implies that for all $\Delta(l_A,l_B)$,

$$\tau_2(\Delta)<\tau_3(\Delta)<\tau_4(\Delta).$$

Inspection of Equation S.7 implies that—all else equal—Δ *increases* in $\tilde{H}(4;l_A,l_B)$; likewise, it *decreases* in $\tilde{H}(3;l_A,l_B)$ and in $\tilde{H}(2;l_A,l_B)$.

First, we show that $l_A(2)=3$. Suppose not, that is $l_A(2)>3$, and let $i>2$ denote the rank of 3, so $l_A(i)=3$. There are three cases: (i) $i=3$, (ii) $i=4$ and $l_A(3)=4$, (iii) $i=4$ and $l_A(3)=5$.

Case 1: $i=3$. Consider the list l'_A that sets $l'_A(2)=3$ and $l'_A(i)=l_A(2)$ and otherwise replicates l_A. Then, $\tilde{H}(3;l_A,l_B)=\tilde{H}(3;l'_A,l_B)$ and $\tilde{H}(3;l_A,l_B)=\tilde{H}(3;l'_A,l_B)$; however:

$$\begin{aligned}&\tilde{H}(2;l'_A,l_B)-\tilde{H}(2;l_A,l_B)\\&\qquad=H\left(l_A(1)+l_A(2)+\underbrace{3-l_A(2)}_{<0}+\sum_{j=1}^{3}l_B(j),\max_{j\in\{1,2,3\}}l_B(j)\right)\\&\qquad-H\left(l_A(1)+l_A(2)+\sum_{j=1}^{3}l_B(j),\max_{j\in\{1,2,3\}}l_B(j)\right)<0.\end{aligned}$$

This implies that l'_A is strictly preferred to l_A for any l_B, i.e., that l_A is strictly dominated.

Case 2: $i=4$ and $l_A(3)=4$. This also implies that $l_A(2)=5$. Consider the list l'_A that sets $l'_A(2)=4$ and $l'_A(3)=5$ and otherwise replicates l_A. Then, $\tilde{H}(3;l_A,l_B)=\tilde{H}(3;l'_A,l_B)$ and $\tilde{H}(4;l_A,l_B)=\tilde{H}(4;l'_A,l_B)$; however:

$$\begin{aligned}&\tilde{H}(2;l'_A,l_B)-\tilde{H}(2;l_A,l_B)\\&=H\left(l_A(1)+5+\underbrace{4-5}_{<0}+\sum_{j=1}^{3}l_B(j),\max_{j\in\{1,2,3\}}l_B(j)\right)-H\left(l_A(1)+5+\sum_{j=1}^{3}l_B(j),\max_{j\in\{1,2,3\}}l_B(j)\right)<0.\end{aligned}$$

This implies that l'_A is strictly preferred to l_A for any l_B, i.e., that l_A is strictly dominated.

Case 3: $i = 4$ and $l_A(3) = 5$. This implies that $l_A(2) = 4$. Consider the list l'_A that sets $l'_A(2) = 3$ and $l'_A(4) = 4$ and otherwise replicates l_A. In terms of induced $\tilde{H}(k; \cdot, l_B)$, the two lists differ in terms of both $\tilde{H}(2; l_A, \cdot)$ and $\tilde{H}(3; l_A, \cdot)$. It is immediate to see that

$$\begin{aligned}&\tilde{H}(2; l'_A, l_B) - \tilde{H}(2; l_A, l_B)\\ &= H\left(l_A(1) + 4 + \underbrace{3-4}_{<0} + \sum_{j=1}^{3} l_B(j), \max_{j\in\{1,2,3\}} l_B(j)\right) - H\left(l_A(1) + 4 + \sum_{j=1}^{3} l_B(j), \max_{j\in\{1,2,3\}} l_B(j)\right) < 0\end{aligned}$$

and

$$\begin{aligned}&\tilde{H}(3; l'_A, l_B) - \tilde{H}(3; l_A, l_B)\\ &= H\left(l_A(1) + 4 + 5 + \underbrace{3-4}_{<0} + \sum_{j=1}^{2} l_B(j), 5\right) - H\left(l_A(1) + 4 + 5 + \sum_{j=1}^{2} l_B(j), 5\right) < 0\end{aligned}$$

This implies that l'_A is strictly preferred to l_A for any l_B, i.e., that l_A is strictly dominated.

Having established that $l_A(2) = 3$, we just need to show that $l_A(3) = 4$. Suppose not, that is $l_A(3) = 5$, and consider the list l'_A that sets $l'_A(3) = 4$ and $l'_A(4) = 5$ and otherwise replicates l_A. Then, $\tilde{H}(2; l_A, l_B) = \tilde{H}(2; l'_A, l_B)$ and $\tilde{H}(4; l_A, l_B) = \tilde{H}(4; l'_A, l_B)$; however:

$$\begin{aligned}&\tilde{H}(3; l'_A, l_B) - \tilde{H}(3; l_A, l_B)\\ &= H\left(l_A(1) + 3 + 5 + \underbrace{4-5}_{<0} + \sum_{j=1}^{2} l_B(j), 5 + \underbrace{4-5}_{<0}\right) - H\left(l_A(1) + 3 + 5 + \sum_{j=1}^{2} l_B(j), 5\right) < 0\end{aligned}$$

This implies that l'_A is strictly preferred to l_A for any l_B, i.e., that l_A is strictly dominated. The argument that we must have $l_B(2) = 5$, $l_B(3) = 4$, $l_B(4) = 3$ is similar.

□

For our next result, define l^*:

$$l^*_J(1) = 5,\ l^*_J(2) = 4,\ l^*_J(3) = 3,\ l^*_J(4) = 2,\ l^*_J(5) = 1.$$

Proposition 2 *If Condition 1 is satisfied, and θ is not too large*:

1. *for any list l_{-J}, l^* is party J's unique best response;*

2. *if Condition 2 is also satisfied, then for any $l' \neq l^*$:*

$$\Pi_A(l^*, l^*) - \Pi_A(l', l^*) > \Pi_B(l^*, l^*) - \Pi_B(l', l^*). \tag{S.8}$$

Proof of Proposition 2. *Part (i)* Let

$$p_J(k|l_J, l_{-J}) \equiv \Pr(\pi(k) < \pi_J < \pi(k+1)).$$

Consider the relative value to party J from employing list l^*, rather than a list $l' \in L^*$, when party $-J$ chooses list l_{-J}:

$$\begin{aligned}\Pi_J(l^*, l_{-J}) - \Pi_J(l', l_{-J}) &= \sum_{k=1}^{4} p_J(k|l^*, l_{-J})\tilde{G}(k; l^*, l_{-J}) - \sum_{k=1}^{4} p_J(k|l', l_{-J})\tilde{G}(k; l', l_{-J}) \\ &= \sum_{k=1}^{4}[p_J(k|l^*, l_{-J})(\tilde{G}(k; l^*, l_{-J}) - \tilde{G}(k; l', l_{-J})) + R_J^{\theta}(k; l^*, l', l_{-J})\tilde{G}(k; l', l_{-J})], \end{aligned} \tag{S.9}$$

where

$$R_J^{\theta}(k; l^*, l', l_{-J}) = -p_J(k|l', l_{-J}) + p_J(k|l^*, l_{-J}). \tag{S.10}$$

Suppose that $\Delta(l^*, l_{-J}) \geq \Delta(l', l_{-J})$. Then $R_J^{\theta}(k; l^*, l', l_{-J}) \geq 0$, and since for each $k \in \{1, \ldots, 4\}$, $\tilde{G}(k; l^*, l_{-J}) - \tilde{G}(k; l', l_{-J}) \geq 0$, l^* is optimal.

Suppose, instead, $\Delta(l^*, l_{-J}) < \Delta(l', l_{-J})$. Notice that $p_J(k|l_J, l_{-J}) = p_{-J}(5-k|l_J, l_{-J})$ and recall that $\tilde{\pi}(3) = 0$ and

$$\tilde{\pi}(2) = -\tilde{\pi}(4) \equiv \tilde{\pi} < 0.$$

Let $\Delta^0(l_A, l_B) = \lim_{\theta \to 0} \Delta(l_A, l_B)$. We have

$$\lim_{\theta \to 0} p_A(1|l_A, l_B) = \frac{1}{2} + \psi\tilde{\pi} - \psi\Delta^0(l_A, l_B) \tag{S.11}$$

$$\lim_{\theta \to 0} p_A(2|l_A, l_B) = -\psi\tilde{\pi} \tag{S.12}$$

$$\lim_{\theta \to 0} p_A(3|l_A, l_B) = -\psi\tilde{\pi} \tag{S.13}$$

$$\lim_{\theta \to 0} p_A(4|l_A, l_B) = \frac{1}{2} + \psi\tilde{\pi} + \psi\Delta^0(l_A, l_B). \tag{S.14}$$

Hence

$$\begin{aligned}\lim_{\theta \to 0} \Big\{\Pi_J(l^*, l_{-J}) - \Pi_J(l', l_{-J})\Big\} = \Big[\frac{1}{2} + \psi\tilde{\pi}\Big]\Big[\tilde{G}(4; l^*, l_{-J}) - \tilde{G}(4; l', l_{-J}) + \tilde{G}(1; l^*, l_{-J}) - \tilde{G}(1; l', l_{-J})\Big] \\ -\psi\tilde{\pi}\Big[\tilde{G}(3; l^*, l_{-J}) - \tilde{G}(3; l', l_{-J}) + \tilde{G}(2; l^*, l_{-J}) - \tilde{G}(2; l', l_{-J})\Big]\end{aligned}$$

$$+\,\psi\Delta^0(l^*,l_{-J})\Big[\tilde{G}(4;l^*,l_{-J})-\tilde{G}(1;l^*,l_{-J})\Big]-\psi\Delta^0(l',l_{-J})\Big[\tilde{G}(4;l',l_{-J})-\tilde{G}(1;l',l_{-J}).\Big]$$

Suppose that $l'(1)=l^*(1)=5$. As θ approaches zero, only the aggregate quality of the candidates in the contested ranks matter for Δ:

$$\Delta^0(l_J,l_{-J})=\frac{1}{3}\sum_{k=2}^{4}\Big\{\tilde{H}(k,l_J,l_{-J})-\tilde{H}(k-1,l_J,l_{-J})\Big\}$$

As a consequence,

$$\begin{aligned}\Delta^0(l',l_{-J})-\Delta^0(l^*,l_{-J})&\propto\tilde{H}(4,l',l_{-J})-\tilde{H}(1,l',l_{-J})-\tilde{H}(4,l^*,l_{-J})+\tilde{H}(1,l^*,l_{-J})\\&=\tilde{H}(4,l',l_{-J})-\tilde{H}(4,l^*,l_{-J})\le 0,\end{aligned}$$

which is a contradiction. Hence, we must have $l'(1)\neq 5$, which in turn implies that $\tilde{G}(1;l^*,l_{-J})-\tilde{G}(1;l',l_{-J})>0$.

Moreover, we can restrict l' to (i) $l'(5)=1$ and (ii) l' strictly decreases in candidate quality in the contested ranks. The reason is that for any list l'' that does not satisfy either restrictions, we can find a list l''' that is preferred by party J to l'', i.e., yields weakly higher Δ^0 and weakly higher $\tilde{G}$, with one strict inequality:

$$\lim_{\theta\to 0}\Big\{\Pi_J(l''',l_{-J})-\Pi_J(l'',l_{-J})\Big\}>0.$$

Since $l'(5)=1$, l^* and l' yield the same set of elected councilors when party A receives four seats:

$$\tilde{H}(4,l^*,l_{-J})=\tilde{H}(4,l',l_{-J})$$

Combining the latter fact with the definition of Δ^0 yields

$$\begin{aligned}\Delta^0(l',l_{-J})-\Delta^0(l^*,l_{-J})&=\frac{1}{3}\Big[\tilde{H}(1,l^*,l_{-J})-\tilde{H}(1,l',l_{-J})\Big]\\&=\frac{1}{3}\left[H\left(5+\sum_{i=1}^{4}l_{-J}(i),\max_{i<5}l_{-J}(i)\right)-H\left(l'(1)+\sum_{i=1}^{4}l_{-J}(i),\max_{i<5}l_{-J}(i)\right)\right]\\&\le\frac{1}{3}\left[H\left(5+14,5\right)-H\left(2+14,5\right)\right]=\frac{1}{3}\int_{16}^{19}H_1(z,5)\,dz\le\frac{1}{3}\kappa\int_{16}^{19}dz=\kappa.\end{aligned}$$

As a consequence—and using again $\tilde{G}(4;l^*,l_{-J})=\tilde{G}(4;l',l_{-J})$—we can write

$$\lim_{\theta\to 0}\Big\{\Pi_J(l^*,l_{-J})-\Pi_J(l',l_{-J})\Big\}\ge\Big[\frac{1}{2}+\psi\tilde{\pi}\Big]\Big[\tilde{G}(1;l^*,l_{-J})-\tilde{G}(1;l',l_{-J})\Big]$$

$$+\,\psi\Delta^0(l^*,l_{-J})\Big[\tilde{G}(4;l^*,l_{-J})-\tilde{G}(1;l^*,l_{-J})\Big]$$
$$-\,\psi\Big[\Delta^0(l^*,l_{-J})+\kappa\Big]\Big[\tilde{G}(4;l^*,l_{-J})-\tilde{G}(1;l',l_{-J})\Big]$$
$$=\Big[\frac{1}{2}+\psi\tilde{\pi}-\psi\Delta^0(l^*,l_{-J})\Big]\Big[\tilde{G}(1;l^*,l_{-J})-\tilde{G}(1;l',l_{-J})\Big]$$
$$-\,\psi\kappa\Big[\tilde{G}(4;l^*,l_{-J})-\tilde{G}(1;l',l_{-J})\Big].$$

Since the distribution of mayoral assignments under (l^*,l_{-J}) and (l',l_{-J}) does not change, Condition 1 implies that κ and $\Delta^0(l^*,l_{-J})$ are small enough relative to the term $\tilde{G}(1;l^*,l_{-J})-\tilde{G}(1;l',l_{-J})>0$. This implies that for each list l_{-J} there exists a finite number $\kappa(l_{-J})$ such that

$$\frac{\kappa(l_{-J})}{\frac{1}{2}+\psi\tilde{\pi}-\psi\Delta^0(l^*,l_{-J})}=\frac{\tilde{G}(1;l^*,l_{-J})-\tilde{G}(1;l',l_{-J})}{\tilde{G}(4;l^*,l_{-J})-\tilde{G}(1;l',l_{-J})}$$

we can then define $\kappa\equiv\min_{l_{-J}\in L(Q)}\kappa(l_{-J})$.

Part (ii) Consider a potential deviation l' that for all θ yields a higher winning probability to J. By a reasoning similar to the one in the previous part of this proof, it must be that $l'(1)\neq 5$, $l'(5)=1$ and that l' is strictly decreasing in the contested ranks.
To prove the second claim, recall that $p_B(k|l,l')=p_A(5-k|l,l')$ and consider the difference:

$$\begin{aligned}
&\Pi_A(l^*,l^*)-\Pi_A(l',l^*)-(\Pi_B(l^*,l^*)-\Pi_B(l',l^*))\\
&=\sum_{k=1}^{4}[p_A(k|l^*,l^*)(\tilde{G}(k;l^*,l_{-J})-\tilde{G}(k;l',l_{-J}))+R^\theta_A(k;l^*,l',l^*)\tilde{G}(k;l',l^*)]\\
&-\sum_{k=1}^{4}[p_B(k|l^*,l^*)(\tilde{G}(k;l^*,l_{-J})-\tilde{G}(k;l',l_{-J}))+R^\theta_B(k;l^*,l',l^*)\tilde{G}(k;l',l^*)]\\
&=\sum_{k=1}^{4}[p_A(k|l^*,l^*)-p_B(k|l^*,l^*)](\tilde{G}(k;l^*,l_{-J})-\tilde{G}(k;l',l_{-J}))\\
&+\sum_{k=1}^{4}[R^\theta_A(k;l^*,l',l^*)-R^\theta_B(k;l^*,l',l^*)]\tilde{G}(k;l',l^*)\\
&=\sum_{k=1}^{4}[p_A(k|l^*,l^*)-p_A(5-k|l^*,l^*)](\tilde{G}(k;l^*,l_{-J})-\tilde{G}(k;l',l_{-J}))\\
&+\sum_{k=1}^{4}[R^\theta_A(k;l^*,l',l^*)-R^\theta_B(k;l^*,l',l^*)]\tilde{G}(k;l',l^*).
\end{aligned}\tag{S.15}$$

We therefore obtain:

$$
\begin{aligned}
&\Pi_A(l^*,l^*)-\Pi_A(l',l^*)-(\Pi_B(l^*,l^*)-\Pi_B(l',l^*))\\
&=p_A(4|l^*,l^*)-p_A(1|l^*,l^*)[\tilde{G}(4|l^*,l^*)-\tilde{G}(4|l',l^*)-(\tilde{G}(1|l^*,l^*)-\tilde{G}(1|l',l^*))]\\
&+p_A(3|l^*,l^*)-p_A(2|l^*,l^*)[\tilde{G}(3|l^*,l^*)-\tilde{G}(3|l',l^*)-(\tilde{G}(2|l^*,l^*)-\tilde{G}(2|l',l^*))]\\
&+\sum_{k=1}^{4}[R_A^\theta(k;l^*,l',l^*)-R_B^\theta(k;l^*,l',l^*)]\tilde{G}(k;l',l^*).
\end{aligned}
\tag{S.16}
$$

We wish to show $\lim_{\theta\to 0}[\Pi_A(l^*,l^*)-\Pi_A(l',l^*)-(\Pi_B(l^*,l^*)-\Pi_B(l',l^*))]>0$. Under Condition 2, we have

$$
\begin{aligned}
\tilde{G}(4|l^*,l^*)-\tilde{G}(4|l',l^*)-(\tilde{G}(1|l^*,l^*)-\tilde{G}(1|l',l^*))>0\\
\tilde{G}(3|l^*,l^*)-\tilde{G}(3|l',l^*)-(\tilde{G}(2|l^*,l^*)-\tilde{G}(2|l',l^*))>0.
\end{aligned}
$$

Let $\Delta^*\equiv\Delta(l^*,l^*)$, $\Delta'_A\equiv\Delta(l',l^*)$, and $\Delta'_B\equiv\Delta(l^*,l')$. We have, since $\Delta^*\xrightarrow[\theta\to 0]{}0$, that

$$
\begin{aligned}
p_A(4|l^*,l^*)-p_A(1|l^*,l^*)&=1-F(-\tilde{\pi}-\Delta^*)-F(\tilde{\pi}-\Delta^*)\xrightarrow[\theta\to 0]{}0\\
p_A(3|l^*,l^*)-p_A(2|l^*,l^*)&=F(-\tilde{\pi}-\Delta^*)+F(\tilde{\pi}-\Delta^*)-2F(-\Delta^*)\xrightarrow[\theta\to 0]{}0
\end{aligned}
$$

We argue that both of these terms are strictly positive, for all $\theta>0$. To see this, note first that for θ sufficiently small, $-\Delta^*<\bar{\xi}$. Then, observe that $F(\cdot)$ is strictly convex on the interval $[-(2\psi)^{-1},\bar{\xi}]$, and strictly concave on the interval $[\bar{\xi},(2\psi)^{-1}]$, and recall that $\tilde{f}(\cdot)$ is symmetric around its mode $\bar{\xi}$. There are two possible cases.

<u>*Case 1*</u>: Suppose, first, $-\tilde{\pi}-\Delta^*\leq\bar{\xi}$. Then,

$$
\begin{aligned}
1-F(-\tilde{\pi}-\Delta^*)-F(\tilde{\pi}-\Delta^*)>&F(\bar{\xi}+\bar{\xi}+\tilde{\pi}+\Delta^*)-F(\tilde{\pi}-\Delta^*)\\
>&F(\bar{\xi})-F(\tilde{\pi}-\Delta^*)>\frac{1}{2}-F(\tilde{\pi}-\Delta^*)>0
\end{aligned}
$$

so that $p_A(4|l^*,l^*)-p_A(1|l^*,l^*)$ is strictly positive. And, the convexity of $F(x)$ in $x<\bar{\xi}$ further implies

$$
\frac{1}{2}F(-\tilde{\pi}-\Delta^*)+\frac{1}{2}F(\tilde{\pi}-\Delta^*)>F\left(\frac{1}{2}\left(-\tilde{\pi}-\Delta^*\right)+\frac{1}{2}\left(\tilde{\pi}-\Delta^*\right)\right)=F(-\Delta^*),
$$

so that $p_A(3|l^*,l^*)-p_A(2|l^*,l^*)$ is strictly positive.

Case 2: Suppose, instead, that $-\tilde{\pi} - \Delta^* > \overline{\xi}$. Then, since $\overline{\xi} > 0$, we have that $-\tilde{\pi} - \Delta^* - \overline{\xi} < \overline{\xi} - (\tilde{\pi} - \Delta^*)$, and symmetry of $\tilde{f}(\cdot)$ yields

$$1 - F(-\tilde{\pi} - \Delta^*) - F(\tilde{\pi} - \Delta^*) > F(\tilde{\pi} - \Delta^*) - F(\tilde{\pi} - \Delta^*) = 0, \tag{S.17}$$

so that $p_A(4|l^*, l^*) - p_A(1|l^*, l^*)$ is strictly positive. Finally, consider the expression:

$$\Lambda(x) = F(x - \Delta^*) - F(-x - \Delta^*) - 2F(-\Delta^*). \tag{S.18}$$

We observe that $\Lambda'(x) = f(x - \Delta^*) - f'(-x - \Delta^*)$ which is strictly positive for $x > 0$. Since $\Lambda(0) = 0$, we conclude that $\Lambda(-\tilde{\pi}) > 0$ since $\tilde{\pi} = \tilde{\pi}(2) < 0$.

Next, we observe:

$$R_A^\theta(k; l^*, l', l^*) - R_B^\theta(k; l^*, l', l^*) = p_A(k|l^*, l^*) - p_A(5-k|l^*, l^*) - p_A(k|l', l^*) + p_A(5-k|l^*, l')$$
$$= \begin{cases} (1-\theta)\psi(\Delta'_A - \Delta^*) + \theta[\tilde{F}(\tilde{\pi} - \Delta^*) - \tilde{F}(\tilde{\pi} - \Delta'_A)] & \\ -(1-\theta)\psi(\Delta^* - \Delta'_B) - \theta[\tilde{F}(-\tilde{\pi} - \Delta'_B) - \tilde{F}(-\tilde{\pi} - \Delta^*)] & k = 1 \\ \theta[\tilde{F}(-\Delta^*) - \tilde{F}(\tilde{\pi} - \Delta^*) - \tilde{F}(-\Delta'_A) + \tilde{F}(\tilde{\pi} - \Delta'_A)] & \\ -\theta[\tilde{F}(-\tilde{\pi} - \Delta^*) - \tilde{F}(-\Delta^*) - \tilde{F}(-\tilde{\pi} - \Delta'_B) + \tilde{F}(-\Delta'_B)] & k = 2 \\ -\theta[\tilde{F}(-\Delta^*) - \tilde{F}(\tilde{\pi} - \Delta^*) - \tilde{F}(-\Delta'_B) + \tilde{F}(\tilde{\pi} - \Delta'_B)] & \\ +\theta[\tilde{F}(-\tilde{\pi} - \Delta^*) - \tilde{F}(-\Delta^*) - \tilde{F}(-\tilde{\pi} - \Delta'_A) + \tilde{F}(-\Delta'_A)] & k = 3 \\ -(1-\theta)\psi(\Delta'_B - \Delta^*) - \theta[\tilde{F}(\tilde{\pi} - \Delta^*) - \tilde{F}(\tilde{\pi} - \Delta'_B)] & \\ +(1-\theta)\psi(\Delta^* - \Delta'_A) + \theta[\tilde{F}(-\tilde{\pi} - \Delta'_A) - \tilde{F}(-\tilde{\pi} - \Delta^*)]. & k = 4 \end{cases}$$

Since $\lim_{\theta \to 0} \Delta'_A = \lim_{\theta \to 0} \Delta'_B = \lim_{\theta \to 0} \Delta^*$, we obtain $R_A^\theta(k; l^*, l', l^*) - R_B^\theta(k; l^*, l', l^*) \xrightarrow[\theta \to 0]{} 0$ for each $k \in \{1, 2, 3, 4\}$. In light of this, to complete the proof we need to show that for $k \in \{3, 4\}$, $p_A(k|l^*, l^*) - p_A(5-k|l^*, l^*)$ converges to zero more slowly than

$$R_A^\theta(k; l^*, l', l^*) - R_B^\theta(k; l^*, l', l^*)$$

and

$$R_A^\theta(5-k; l^*, l', l^*) - R_B^\theta(5-k; l^*, l', l^*).$$

Now, we argue that (we make the dependence of Δ on θ explicit)

$$\Delta'_A(\theta) + \Delta'_B(\theta) - 2\Delta^*(\theta) = O(\Delta^*(\theta)) \quad \text{as} \quad \theta \to 0.$$

To see this, notice that using L'Hopital's rule and the Implicit Function Theorem, we obtain:

$$\lim_{\theta\to 0}\frac{\Delta'_A(\theta)}{\Delta^*(\theta)}=\lim_{\theta\to 0}\frac{d\Delta'_A(\theta)/d\theta}{d\Delta^*(\theta)/d\theta}=\frac{\sum_{k=2}^{4}[2\tilde{f}(\tilde{\pi}(k))-\sum_{r\neq k}\tilde{f}(\tilde{\pi}(r))][\tilde{H}(k;l',l^*)-\tilde{H}(k-1;l',l^*)]}{\sum_{k=2}^{4}[2\tilde{f}(\tilde{\pi}(k))-\sum_{r\neq k}\tilde{f}(\tilde{\pi}(r))][\tilde{H}(k;l^*,l^*)-\tilde{H}(k-1;l^*,l^*)]}<\infty$$

which follows from

$$\frac{d\Delta(\theta)}{d\theta}=-\frac{\sum_{k=2}^{4}\frac{d\tau_k(\Delta,\theta)}{d\theta}[\tilde{H}(k;l_A,l_B)-\tilde{H}(k-1;l_A,l_B)]}{\sum_{k=2}^{4}\frac{d\tau_k(\Delta,\theta)}{d\Delta}[\tilde{H}(k;l_A,l_B)-\tilde{H}(k-1;l_A,l_B)-1},$$

$\lim_{\theta\to 0}\frac{d\tau_k(\Delta,\theta)}{d\Delta}=0$, and

$$\frac{d\tau_k(\Delta,\theta)}{d\theta}=\frac{3\psi[\tilde{f}(\tilde{\pi}(k))-\psi]-\left(\sum_{r=2}^{4}\{\tilde{f}(\tilde{\pi}(r))\}-3\psi\right)\psi}{9\psi^2}=\frac{2\tilde{f}(\tilde{\pi}(k))-\sum_{r\neq k}\tilde{f}(\tilde{\pi}(r))}{9\psi}.$$

The argument for $\Delta'_B(\theta))$ is analogous. In light of this, and since $\lim_{\theta\to 0}\Delta'_A=\lim_{\theta\to 0}\Delta'_B=\lim_{\theta\to 0}\Delta^*$, for each we can write

$$\begin{aligned}R^{\theta}_A(k;l^*,l',l^*)-R^{\theta}_B(k;l^*,l',l^*)&=O\big(\Delta^*(\theta)+\theta\tilde{D}^k_F(\theta)\big)\quad\text{as}\quad\theta\to 0\quad k\in\{1,4\}\\ R^{\theta}_A(k;l^*,l',l^*)-R^{\theta}_B(k;l^*,l',l^*)&=O\big(\theta\tilde{D}^k_F(\theta)\big)\quad\text{as}\quad\theta\to 0\quad k\in\{2,3\}\end{aligned}$$

where

$$\tilde{D}^k_F(\theta)=\begin{cases}\tilde{F}(\tilde{\pi}-\Delta^*)-\tilde{F}(\tilde{\pi}-\Delta'_A)-\tilde{F}(-\tilde{\pi}-\Delta'_B)+\tilde{F}(-\tilde{\pi}-\Delta^*)] & k=1\\ \tilde{F}(-\Delta^*)-\tilde{F}(\tilde{\pi}-\Delta^*)-\tilde{F}(-\Delta'_A)+\tilde{F}(\tilde{\pi}-\Delta'_A) & \\ -\tilde{F}(-\tilde{\pi}-\Delta^*)+\tilde{F}(-\Delta^*)+\tilde{F}(-\tilde{\pi}-\Delta'_B)-\tilde{F}(-\Delta'_B)] & k=2\\ -\tilde{F}(-\Delta^*)+\tilde{F}(\tilde{\pi}-\Delta^*)+\tilde{F}(-\Delta'_B)-\tilde{F}(\tilde{\pi}-\Delta'_B) & \\ +\tilde{F}(-\tilde{\pi}-\Delta^*)-\tilde{F}(-\Delta^*)-\tilde{F}(-\tilde{\pi}-\Delta'_A)+\tilde{F}(-\Delta'_A)] & k=3\\ -\tilde{F}(\tilde{\pi}-\Delta^*)+\tilde{F}(\tilde{\pi}-\Delta'_B)+\tilde{F}(-\tilde{\pi}-\Delta'_A)-\tilde{F}(-\tilde{\pi}-\Delta^*) & k=4\end{cases}$$

goes to zero as $\theta\to 0$. This implies that there is no positive constant C for which

$$\begin{aligned}R^{\theta}_A(k;l^*,l',l^*)-R^{\theta}_B(k;l^*,l',l^*)&=O\big(\Delta^*(\theta)+\theta C\big)\quad\text{as}\quad\theta\to 0\quad k\in\{1,4\}\\ R^{\theta}_A(k;l^*,l',l^*)-R^{\theta}_B(k;l^*,l',l^*)&=O\big(\theta C\big)\quad\text{as}\quad\theta\to 0\quad k\in\{2,3\}\end{aligned}$$

On the other hand, since for all θ

$$[1 - \tilde{F}(\tilde{\pi} - \Delta^*) - \tilde{F}(-\tilde{\pi} - \Delta^*)] \underset{\theta \to 0}{\longrightarrow} 1 - \tilde{F}(\tilde{\pi}) - \tilde{F}(-\tilde{\pi}) > 0$$

$$[\tilde{F}(\tilde{\pi} - \Delta^*) + \tilde{F}(-\tilde{\pi} - \Delta^*) - 2\tilde{F}(-\Delta^*)] \underset{\theta \to 0}{\longrightarrow} \tilde{F}(\tilde{\pi}) + \tilde{F}(-\tilde{\pi}) - 2\tilde{F}(0) > 0$$

there exists a constant $C > 0$ such that

$$p_A(k|l^*, l^*) - p_A(5 - k|l^*, l^*) = O(\Delta^*(\theta) + \theta C) \quad \text{as} \quad \theta \to 0 \quad k \in \{1, 4\}$$

$$p_A(k|l^*, l^*) - p_A(5 - k|l^*, l^*) = O(\theta C) \quad \text{as} \quad \theta \to 0 \quad k \in \{2, 3\}.$$

This completes the proof. $\square$

Appendix B: Descriptive Statistics and Additional Figure

We first present Descriptive Statistics for our Estimation Sample.

	Leadership Score	Cognitive Score	Earnings Score	Years of Education
Top Ranked	6,22 (1,73)	6,20 (1,67)	0,74 (0,95)	13,6 (3,0)
Safe	5,93 (1,74)	5,90 (1,67)	0,64 (0,88)	12,9 (3,0)
Advantaged Contested	5,76 (1,78)	5,85 (1,80)	0,40 (0,89)	13,0 (2,9)
Highly Contested	5,76 (1,79)	5,69 (1,82)	0,37 (0,91)	12,9 (2,9)
Disadvantaged Contested	5,72 (1,76)	5,69 (1,82)	0,30 (0,92)	13,0 (2,9)
Certain Loss	5,62 (1,75)	5,55 (1,85)	0,23 (0,94)	12,7 (3,0)
All Categoriess	5.72 (1,76)	5.67 (1,76)	0,34 (0,94)	12,8 (3,0)

Figure S.1 – **Descriptive statistics for valence measures across ballot rank categories**. The table shows means and standard deviations for four valence measures within six categories of ballot ranks, and for all categories combined.

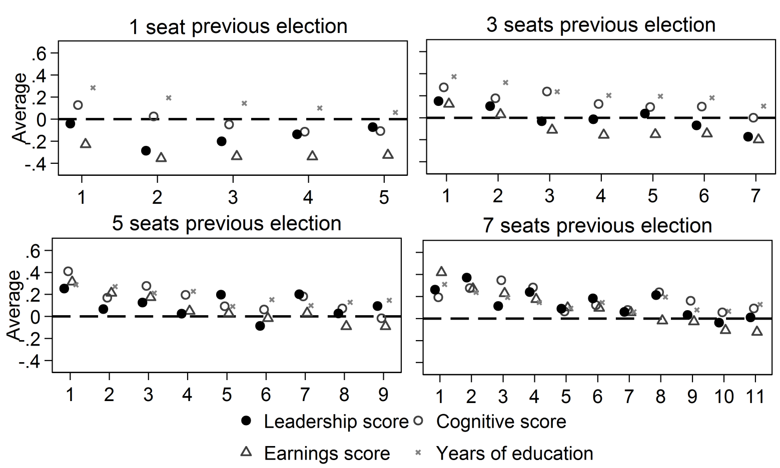

Figure S.2 – **List rank and Valence**. The y-axis shows the average of each of the four valence measurements, transformed into Z-scores. Data includes all party groups between 1991 and 2014 that won the specified odd number of seats in the previous election.

Second, we extend our presentation of Figure 3—which plots the relationship between list rank and valence for parties that won between 1 and 8 seats in the previous election, for even numbers—to odd-numbered seats 1, 3, 5, and 7.

Appendix C: Additional Theoretical Results: Flexible Lists

In this Appendix, we extend our theoretical results to *flexible list* settings, which constitute our post-1998 municipal election data. Under a flexible list system, each voter may cast a "preference" vote for an individual candidate on a party's list. Seats are awarded to parties on the basis of the relative share of preference votes that accrue to candidates on each party's list. The allocation of seats within parties corresponds to the order in which candidates are ranked on the electoral ballot, with one important exception: if a candidate's share of preference votes within her party's list exceeds a pre-specified threshold, she acquires electoral priority over her co-partisans whose share did not exceed the threshold *regardless* of her position on the list. The threshold varies across countries: in our Swedish municipal data, flexible lists are first employed in 1998, and consistently maintain a 5% threshold.

Preference Voting. To capture this context, we extend our benchmark model to flexible lists by assuming that each voter now casts her ballot (a "preference vote") for a single candidate on either party's list. We let π_J^i denote the share of preference votes that accrue to candidate $i \in \{1, ..., 5\}$ on party $J \in \{A, B\}$'s list. Thus, party J's total share of preference votes is

$$\pi_J = \sum_{i \in \{1,...,5\}} \pi_J^i. \tag{S.19}$$

Seats are allocated at the party level according to the same mechanism as our benchmark model: when π_J is between the thresholds $\pi(k)$ and $\pi(k+1)$, the number of seats awarded to party J is $n_J = k$, while party $-J$ is awarded $n_{-J} = 5 - k$ seats.

How Seats are Awarded. Let $P_J = \{i : \frac{\pi_J^i}{\pi_J} \geq \zeta\}$, where $\zeta \in (0, 1]$. We define a function $p_J : P_J \Rightarrow \{1, 2, ..., |P_J|\}$ such that for any two candidates $i, j \in P_J$: if $\pi_i^J > \pi_j^J$, then $p_J(i) < p_J(j)$; and if $\pi_i^J = \pi_j^J$, $p_J(i) < p_J(j)$, if and only if $l_J^{-1}(i) < l_J^{-1}(j)$. The allocation of seats proceeds as follows:

1. if $n_J \leq |P_J|$, candidates $p_J^{-1}(1), p_J^{-1}(2), ..., p_J^{-1}(n_J)$ are awarded seats;
2. if $n_J > |P_J|$, all candidates in P_J are awarded seats; the remaining $n_J - |P_J|$ seats are filled by those candidates that are not contained in P_J, according to the order in which they listed on the ballot.

In words: the set P_J identifies candidates whose individual share of the preference votes on the party's ballot (i.e., $\frac{\pi_J^i}{\pi_J}$) exceeds a pre-specified threshold (ζ)—e.g., $\zeta = \frac{1}{20}$ in the Swedish municipal context. The function p_J ranks these candidates from highest ($p_J(i) = 1$) to lowest

($p_J(i) = |P_J|$), with higher rankings awarded to candidates with higher vote share; if any two candidates amongst those receiving a within-party share of preference votes in excess of ζ win the same vote share, the higher ranking is awarded to the candidate with a higher rank according to the list.

Note: The reason for developing this notation is that—in principle—multiple candidates could surpass the electoral threshold, and the electoral rule must specify which of these candidates will win seats in the event that the number of seats won by the party is fewer than the number of candidates that surpass the threshold.

Voting Behavior. With a continuum of voters, any allocation of preference votes within a party's list is a best response. We assume that a share $1-\varepsilon$ of voters (with $\varepsilon > 0$ arbitrarily small) are instrumental: they choose among parties as in the baseline model (i.e., by computing the instrumental values V_A and V_B) and then vote for the candidate with the highest observed quality within the chosen list. The remaining ε voters—referred to as limited knowledge, or "LK" voters—cast their preference vote for the *capolista* of higher quality: they cast a preference vote for $l_A^{-1}(1)$ with probability one if $l_A(1) > l_B(1)$, cast a preference vote for $l_A^{-1}(1)$ with probability one-half if $l_A(1) = l_B(1)$, and cast a preference vote for $l_B^{-1}(1)$ otherwise. Introducing this arbitrarily small group of LK voters serves to eliminate a trivial equilibrium multiplicity in our flexible list context; it does not affect any of the results in our benchmark model of closed lists.[18]

We then define the following operator. For any $l_J \in L(Q)$, the operator $l^F[l_J]$ yields

1. if $l_J^{-1}(5) = 1$: $l^F[l_J](k) = l_J(k)$ for all $k \in \{1, ..., 5\}$.

2. if $l_J^{-1}(5) \neq 1$: $l^F[l_J](1) = 5$, $l^F[l_J](k) = l_J(k-1)$ for any $k \in \{2, ..., l_J^{-1}(5)\}$, and $l^F[l_J](k) = l_J(k)$ for all $k > l_J^{-1}(5)$.

In words: the operator $l^F[\cdot]$ generates the order in which seats are filled under the list l_J when voters who cast their ballots for party J award their preference vote to the candidate of the highest quality. To understand how this operator works, suppose that party J offers a list $l_J = (3, 2, 5, 1, 4)$, i.e., it ranks candidate 3 at the top, followed by 2, and so on, with candidate 1 ranked in the lowest part of the list. Given voting behavior, and $\zeta < 1$, voters and parties anticipate that if the party wins only one seat, candidate 5 will be elected; if the party wins two seats, candidates 5 and 3 will be elected; and so on. Given the voting behavior described above, the *effective ranking* is $l^F[l_J] = (5, 3, 2, 1, 4)$.

[18] Since in the baseline model parties have strict best responses (up to a measure-zero set of parameters), all results continue to hold when a share ε of voters simply cast a vote for the party with the higher-quality *capolista*.

Notice that the order in which seats are filled under the list l_J replicates the order in which seats are filled in a closed-list setting where party J's list is instead $l^F(l_J)$. However, the presence of LK voters implies that in a flexible-list setting, the party's vote share from lists l_J and $l^F(l_J)$ may differ.

Results. We now extend the results of our benchmark model. Since party objectives following the allocation of seats are unchanged, Lemma 1 continues to hold: the majority party continues to appoint its highest-quality elected official as chief executive. We number the following results to be consistent with the results in our benchmark presentation—i.e., the following Lemma S.2 extends Lemma 2 in our main presentation to flexible lists. Throughout, we use the superscript F to denote quantities under flexible lists.

Compared to the baseline model, a list l_J affects each party's vote tally through the behavior of the limited-knowledge voters, who only respond to the choice of the *capolista*. The remaining voters take into consideration the behavior of the LK-voters when making their voting decision. Hence, the salience of each rank becomes a function of the perceived difference in values Δ as well as the "gap" in the quality of the capolista. Let

$$\gamma(l_A, l_B) = \begin{cases} 1 & l_A(1) > l_B(1) \\ 0 & l_A(1) = l_B(1) \\ -1 & l_A(1) < l_B(1) \end{cases}$$

Party A's vote share is then given by

$$\pi_A^F = (1-\varepsilon)\left(\frac{1}{2} + \phi\Delta + \phi\xi\right) + \varepsilon \begin{cases} 1 & \text{if } \gamma(l_A, l_B) = 1 \\ \frac{1}{2} & \text{if } \gamma(l_A, l_B) = 0 \\ -1 & \text{if } \gamma(l_A, l_B) = -1 \end{cases}$$

Adding and subtracting $\frac{\varepsilon}{2}$, and re-arranging, we obtain:

$$\pi_A^F = \frac{1}{2} + (1-\varepsilon)\phi\Delta + (1-\varepsilon)\phi\xi + \gamma\frac{\varepsilon}{2}$$

Hence,

$$\pi_A^F = \pi(k) \Leftrightarrow \xi = \frac{\pi(k) - \frac{1}{2} - \frac{\gamma\varepsilon}{2}}{\phi(1-\varepsilon)} - \Delta \equiv \check{\pi}(k, \gamma) - \Delta$$

We can then define the salience weight as

$$\tau_k(\Delta, \gamma) \equiv \frac{f\left(\breve{\pi}(k,\gamma) - \Delta\right)}{\sum_{r=2}^{4} f\left(\breve{\pi}(r,\gamma) - \Delta\right)}$$

Notice that (i) $\sum_{r=2}^{4} \breve{\pi}(r,\gamma) = -\frac{3}{2}\gamma\frac{\varepsilon}{\phi(1-\varepsilon)}$ (in the baseline model, we have $\sum_{r=2}^{4} \tilde{\pi}(r) = 0$), (ii) $\sum_{r=2}^{4} \tau_k(\Delta,\gamma) = 1$ for all Δ and γ, and (iii) under the assumptions, $\tau_k(\Delta,\gamma) \approx \tau_k(\Delta)$.

Lemma S.2. Under flexible list PR, for any party lists l_A and l_B, there exists a pair $V_A(l_A, l_B)$ and $V_B(l_A, l_B)$, and attention weights $\tau^F(k) \in (0,1)$ for $\kappa \in \{2,3,4\}$ satisfying $\sum_{k=2}^{4} \tau^F(k) = 1$, such that an instrumental voter's value from party B is:

$$V_A^F(l_A, l_B) = \sum_{k=2}^{4} \tau^F(k)\ \tilde{H}(k, l^F[l_A], l^F[l_B]),$$

while her value from party A is:

$$V_B^F(l_A, l_B) = \sum_{k=2}^{4} \tau^F(k)\ \tilde{H}(k-1, l^F[l_A], l^F[l_B]).$$

Moreover, when θ is small enough, for each pair of lists (l_A, l_B), the difference $\Delta^F(l_A, l_B) = V_A^F(l_A, l_B) - V_B^F(l_A, l_B)$ is unique.

Proof. Recall that, conditional on voting for a party's list, instrumental voters cast their preference votes for the candidate with the highest observed quality. For any threshold $\zeta \in (0,1)$, this implies that for any list l_J, the order in which seats are assigned is given by the strict order $l^F[l_J]$. The remainder of the proof is precisely as in our benchmark model, with appropriate substitutions of (l_A, l_B) to $(l^F[l_A], l^F[l_B])$ and $\tau_k(\Delta)$ with $\tau_k(\Delta,\gamma)$ (with $\tau_k(\Delta,\gamma) \approx \tau_k(\Delta)$). □

The intuition is simple: under flexible lists, voters form their values from supporting party A versus party B just as under closed lists, however they recognize that whenever a list l_J is offered by party J, the "effective" list is $l^F(l_J)$. Otherwise, the argument is exactly the same as in our benchmark setting. Lemma 2 implies that, since ε is arbitrarily small, switching from l_J to $l^F[l_J]$ has a negligible effect on Δ and on party J's expected payoff. As a consequence, the payoff difference $\Delta^F(l^F[l_A], l^F[l_B])$ is arbitrarily close to $\Delta(l^F[l_A], l^F[l_B])$, the difference under closed lists.

Corollary 1. *Under the assumptions:*
(i) $\Delta^F(l_A, l_B) \approx \Delta^F(l^F[l_A], l_B) \approx \Delta(l^F[l_A], l_B)$ *and* $\Delta^F(l_A, l_B) \approx \Delta^F(l_A, l^F[l_B]) \approx \Delta(l_A, l^F[l_B])$*;*
(ii) $\Pi_A^F(l_A, l_B) \approx \Pi_A^F(l^F[l_A], l_B)$ *and* $\Pi_B^F(l_A, l_B) \approx \Pi_B^F(l_A, l^F[l_B])$*;*

(iii) when parties are purely seat-motivated, $\Delta^F(l_A, l_B)$ is a sufficient statistic of party A's expected payoff and $-\Delta^F(l_A, l_B)$ is a sufficient statistic of party B's expected payoff.

We now prove a key result: for every list l, a party cannot worsen its expected payoff by switching to $l^F[l]$.

Lemma S.3. *When θ is small enough, for all $l_A \in L(Q)$ and $J \in \{A, B\}$, $l^F[l_A] \succeq_J l_A$.*

Proof. We prove the result for party A. The argument for party B is analogous. First, we show that party A's expected number of seats weakly increases by choice of $l^F[l_A]$, versus l_A, regardless of party B's list. The expected number of seats is:

$$4 - \sum_{k=2}^{4} \Pr(\pi_A \leq \pi(k)) = 4 - \sum_{k=2}^{4} F\Big(\breve{\pi}(k, \gamma(l_A, l_B)) - \Delta^F(l_A, l_B)\Big)$$

There are two cases to consider: if $\gamma(l_A, l_B) = \gamma(l^F[l_A], l_B)$ (that is, the preferences of the LK voters do not change as a result of the switch), then it is immediate to see that, from Lemma 2, $\Delta^F(l_A, l_B) = \Delta^F(l^F[l_A], l_B)$ and thus the expected number of seats is constant across both lists l_A and $l^F[l_A]$. Suppose then that $\gamma(l_A, l_B) \neq \gamma(l^F[l_A], l_B)$. Since the capolista under $l^F[l_A]$ is of the highest quality ($l^F[l_A](1) = 5$), it must be that switching from l_A to $l^F[l_A]$ increases the number of LK voters who prefer party A: $\gamma(l^F[l_A], l_B) > \gamma(l_A, l_B)$. We now show that when θ is small enough, for all $(l_A, l_B) \in L^2(Q)$

$$\breve{\pi}(k, \gamma(l_A, l_B)) - \Delta^F(l_A, l_B) > \breve{\pi}(k, \gamma(l^F[l_A], l_B)) - \Delta^F(l^F[l_A], l_B).$$

To do that, we argue that the difference

$$\mathcal{D}(\varepsilon; l_A, l_B) \equiv \breve{\pi}(k, \gamma(l_A, l_B)) - \breve{\pi}(k, \gamma(l^F[l_A], l_B)) - \Big[\Delta^F(l_A, l_B) - \Delta^F(l^F[l_A], l_B)\Big]$$

(i) approaches zero as $\varepsilon = 0$ and (ii) is increasing in ε when θ is small enough. To see why (ii) is true, rewrite $\mathcal{D}$ as

$$\mathcal{D}(\varepsilon; l_A, l_B) = [\gamma(l^F[l_A], l_B) - \gamma(l_A, l_B)] \frac{\varepsilon}{1-\varepsilon} \frac{1}{2\phi} - \Delta^F(l_A, l_B) + \Delta^F(l^F[l_A], l_B)$$

using the fact that

$$\breve{\pi}(k, \gamma) = \frac{\pi(k) - \frac{1}{2}}{\phi(1-\varepsilon)} - \gamma \frac{\varepsilon}{1-\varepsilon} \frac{1}{2\phi}$$

To complete the proof, notice that for every $l \in L(Q)$ $\Delta^F(l, l_B)$ is the unique fixed point of the mapping

$$\mathcal{V}^F(\Delta,\gamma)=\sum_{k=2}^{4}\tau_k(\Delta,\gamma)\Big[\tilde{H}(k,l^F[l],l^F[l_B])-\tilde{H}(k-1,l^F[l],l^F[l_B])\Big].$$

Hence, it depends on ε via the salience weights $\tau_k(\Delta,\gamma)=\frac{f(\tilde{\pi}(k,\gamma)-\Delta)}{\sum_{r=2}^{4}f(\tilde{\pi}(r,\gamma)-\Delta)}$. Since $\lim_{\theta\to 0}\tau_k(\Delta,\gamma)=\frac{1}{3}$, we must have $\lim_{\theta\to 0}\frac{d\Delta}{d\varepsilon}=0$. This establishes that party A's expected number of seats strictly increases by choice of $l^F[l_A]$.

Next, we argue that $l^F[l_A]$ is weakly preferred by party A to l_A if party A's expected number of seats is weakly higher under the former list than the latter, for any payoff function $G(k,m,\overline{q}_A)$ satisfying Assumption 2. This is immediate from the following observations: (1) under $(l^F[l_A],l_B)$, party A's number of won seats k is weakly higher than under l_A,l_B, and (2) for any post-election number of seats won by party A, the set of elected legislators $S_A\cup S_B$ is the same under both lists. $\square$

Proposition S.1. *Proposition 1 extends to flexible list PR.*

Proof. We proceed by way of four steps.

Step 1. For any l_{-J}, if l_J is a best response then the list $l^F(l_J)$ is also a best response to l_{-J}. Proof. We make the argument for party $J=A$, since the argument for $J=B$ is symmetric. Suppose that l_A is a best response to a list l_B. Since parties are purely seat-motivated, by Corollary 1.iii, this is equivalent to $\Delta^F(l_A,l_B)\geq\Delta^F(l',l_B)$ for all $l'\in L(Q)$. However, Lemma S.3 shows that $\Delta^F(l^F[l_A],l_B)\geq\Delta^F(l_A,l_B)$. $\square$

So, if a pair (l_A,l_B) is an equilibrium under flexible lists when parties are purely seat-motivated, then the pair $(l^F(l_A),l^F(l_B))$ is also an equilibrium under flexible lists when parties are purely seat-motivated. Henceforth, we define $L^F(Q)=\{l\in L(Q):[l^F(l_J)](k)=l_J(k),k\in\{1,...,5\}\}$. That is, $L^F(Q)$ identifies party lists l_J that are unaltered by the operation $l^F(l_J)$. Lemma S.3 implies that if l_J is individually rational for J, there are two possibilities. First, $l_J\in L^F(Q)$. Second, $l_J\notin L^F(Q)$, but switching from l_J to $l^F[l_J]$ does not affect the behavior of the LK voters. In the latter case, we have $\text{sign}\{l_J(1)-l_{-J}(1)\}=\text{sign}\{l^F[l_J](1)-l_{-J}(1)\}$. Since $l^F[l_J](1)=5$ and, by $l_J\notin L^F(Q)$, $l_J(1)<5$, this further requires $l_{-J}(1)\leq 4$.

Step 2. Consider a profile (l_J,l_{-J}). Either $(l_J,l_{-J})\in[L^F(Q)]^2$, or one party has a profitable deviation. Proof. Suppose $l_A\notin L^F(Q)$. If $l_B(1)=5$, then $l^F[l_A]$ is a profitable deviation, since $\gamma(l_A,l_B)=-1$ and $\gamma(l^F[l_A],l_B)=0$. If $l_B(1)\neq 5$, then there are two possible cases: either $\gamma(l_A,l_B)=-1$, in which case $l^F[l_A]$ is again a profitable deviation, or $\gamma(l_A,l_B)\in\{0,1\}$, in which case $l^F[l_B]$ is a profitable deviation for B, since $\gamma(l_A,l^F[_B])=-1$. $\square$

Step 3. A list $l_J\in L^F(Q)$ is a weak best response to at least one list l_{-J} only if $l_J(5)=1$. Proof.

We prove the argument for party A; the argument for party B is the same. Using Lemma 2 and the same arguments as in the proof of Proposition 1—recalling that by Corollary 1).i, $\Delta^F \approx \Delta$—there exists θ small enough such that:

$$\Delta^F(l_A, l_B) > 0 \iff \tilde{H}(4; l^F[l_A], l^F[l_B]) - \tilde{H}(1; l^F[l_A], l^F[l_B]) > 0. \tag{S.20}$$

For any pair (l_A, l_B), under flexible lists:

$$\begin{aligned}&\tilde{H}(4; l^F[l_A], l^F[l_B]) - \tilde{H}(1; l^F[l_A], l^F[l_B]) \\ &= H\left(\sum_{i=1}^{4} l^F[l_A](i) + l^F[l_B](1), 5\right) - H\left(l^F[l_A](1) + \sum_{i=1}^{4} l^F[l_A](i), 5\right).\end{aligned} \tag{S.21}$$

Suppose a list l_A is chosen satisfying $l_A(1) = 5$ (so $l_A \in L^F(Q)$) and $l_A(5) \neq 1$. Let $i = l_A^{-1}(1)$. Consider another list that sets $l'_A(5) = 1$ and $l'_A(i) = l_A(5)$, but otherwise replicates l_A. We obtain that:

$$\begin{aligned}&\tilde{H}(4; l^F[l'_A], l^F[l_B]) - \tilde{H}(1; l^F[l'_A], l^F[l_B]) - \left[\tilde{H}(4; l^F[l_A], l^F[l_B]) - \tilde{H}(1; l^F[l_A], l^F[l_B])\right] \\ =&H\left(\sum_{j=1}^{4} l^F[l_A](j) + \underbrace{l^F[l_A](5) - l^F[l_A](i)}_{>0} + l^F[l_B](1), 5\right) - H\left(\sum_{j=1}^{4} l^F[l_A](j) + l^F[l_B](1), 5\right) > 0,\end{aligned} \tag{S.22}$$

for any list l_B. We conclude that l_A is strictly dominated by l'_A, for θ small. $\square$

Step 4. A list $l_A \in L^F(Q)$ is a weak best response to at least one $l_B \in L^F(Q)$ only if $l_A^{-1}(4) \in \{3, 4\}$. And, a list $l_B \in L^F(Q)$ is a weak best response to at least one $l_A \in L^F(Q)$ only if $l_B^{-1}(4) \in \{2, 3\}$.

Proof. We prove the argument for party A. By the previous steps, we must have $l_A(1) = 5$ (since $l_A \in L^F(Q)$) and $l_A(5) = 1$. Suppose that a list satisfying $l_A(2) = 4$ is a best response to at least one list $l_B \in L^F(Q)$. Consider another list l'_A that sets $l'_A(2) = l_A(3)$ and $l'_A(3) = 4$ and otherwise satisfies $l_A(k) = l'_A(k)$. It is immediate to see that for any list l_B and rank $k \in \{3, 4\}$, $\tilde{H}(k; l^F(l_A), l^F(l_B)) = \tilde{H}(k; l^F(l'_A), l^F_B)$. First, we observe that:

$$\begin{aligned}&\tilde{H}(2; l^F(l'_A), l^F(l_B)) - \tilde{H}(2; l^F(l_A), l^F(l_B)) \\ =&H\left(5 + 4 + \underbrace{[l^F(l_A)](3) - 4}_{<0} + \sum_{j=1}^{3} [l^F(l_B)](j), 5\right) - H\left(5 + 4 + \sum_{j=1}^{3} [l^F(l_B)](j), 5\right) < 0.\end{aligned}$$

Second, we show that $\Delta^F(l^F(l_A), l^F(l_B))$ strictly decreases in $\tilde{H}(2; l^F(l_A), l^F(l_B))$, which implies that $\Delta^F(l^F(l'_A), l^F(l_B)) > \Delta^F(l^F(l_A), l^F(l_B))$. Since going from l_A to l'_A does not affect the behavior of LK voters—$\gamma(l_A, l_B) = \gamma(l'_A, l_B)$—, this completes the argument. By assumption, $(l_A l_B) \in [L^F(Q)]^2$, hence $\tilde{H}(4; l_A, l_B) = \tilde{H}(1; l_A, l_B)$. Recalling that

$$\mathcal{V}^F(\Delta, \gamma) = \sum_{k=2}^{4} \tau_k(\Delta, \gamma)\big[\tilde{H}(k; l^F(l_A), l^F(l_B)) - \tilde{H}(k-1; l^F(l_A), l^F(l_B))\big],$$

we can rewrite $\mathcal{V}^F(\Delta, \gamma)$ as

$$\begin{aligned} \mathcal{V}^F(\Delta, \gamma) = &\tilde{H}(4; l^F(l_A), l^F(l_B))[\tau_4(\Delta, \gamma) - \tau_2(\Delta, \gamma)] \\ &+ \tilde{H}(3; l^F(l_A), l^F(l_B))[\tau_3(\Delta, \gamma) - \tau_4(\Delta, \gamma)] \\ &+ \tilde{H}(2; l^F(l_A), l^F(l_B))[\tau_2(\Delta, \gamma) - \tau_3(\Delta, \gamma)]. \end{aligned} \tag{S.23}$$

Hence, by the implicit function theorem,

$$\frac{d\Delta^F(l^F(l_A), l^F(l_B))}{d\tilde{H}(2; l^F(l_A), l^F(l_B))} = \frac{\tau_2(\Delta, \gamma) - \tau_3(\Delta, \gamma)}{1 - \frac{\partial \mathcal{V}^F}{\partial \Delta}} < 0.$$

The last inequality follows from

$$\lim_{\varepsilon \to 0} \Big\{ \tau_2(\Delta, \gamma) - \tau_3(\Delta, \gamma) \Big\} = \tau_2(\Delta) - \tau_3(\Delta) < 0$$

(which we established earlier) and the the fact that, by the same argument of Lemma 2, $\frac{\partial \mathcal{V}^F}{\partial \Delta} < 1$. The argument for why $l_B(4) \in \{2, 3\}$ follows a similar logic. □

We have shown that a pair (l_A, l_B) is an equilibrium only if (1) $l_A = l^F[l_A]$ and $l_B = l^F[l_B]$, that is $l^F[l_J](1) = 5$ for $J \in \{A, B\}$,(2) $l^F[l_J](5) = 1$, and (3) $l^F[l_A](2) < \max\{l^F[l_A](3), l^F[l_A](4)\}$ and only if $l^F[l_B](4) < \max\{l^F[l_B](2), l^F[l_B](3)\}$. As a consequence, the Marginal Rank Hypothesis continues to hold. Contrary to the closed list setting, however, both parties will place their best candidates at the top of the ballot under flexible lists.

We now argue that Proposition 2 also extends to this setting. An intuition is that under flexible list PR, (1) there exists an *electoral* incentive to place one's best candidate to the top of the ballot that augments intrinsic incentives identified in Proposition 2, and (2) the presence of the flexible list and the small number of LK voters, however, do not affect parties' incentives within the contested ranks, which are as in the baseline model.

Proposition S.2. *Proposition 2 extends to flexible list PR.*

Proof. The proof mirrors the proof of Proposition 2, with minor modifications. Below, we sketch the main steps of the proof and list the minimal modifications required to adapt the argument. The proof is divided in two parts:

Part (i)

(A) Argue that if $\Delta(l^*, l_{-J}) \geq \Delta(l', l_{-J})$, for any list l', then l^* is optimal for J.

The argument under flexible lists is almost unchanged. Replacing $p_J(k, l_J, l_{-J})$ with $p_J^F(k, l_J, l_{-J}) = \Pr(\pi(k) < \pi_J^F < \pi(k+1))$ and $\Pi_J(l_J, l_{-J})$ with $\Pi_J^F(l_J, l_{-J})$, the term R_J^θ becomes

$$R_J^{F\theta}(k; l^*, l', l_{-J}) \equiv -p_J^F(k|l', l_{-J}) + p_J^F(k|l^*, l_{-J}).$$

Since $l^*(1) = 5$, the behavior of the LK voters is such that if $\Delta^F(l^*, l_{-J}) \geq \Delta^F(l', l_{-J})$, we still have $R_J^{F\theta}(k; l^*, l', l_{-J}) \geq 0$.

(B) Assuming $\Delta(l^*, l_{-J}) < \Delta(l^*, l_{-J})$, re-express the difference

$$\lim_{\theta \to 0} \Pi_J(l^*, l_{-J}) - \Pi_J(l', l_{-J})$$

in a way that leads to conclude that (i) $l'(1) \neq 5$, (ii) $l'(5) = 1$ and (iii) $l'(k) < l'(k+1)$ for ranks $k \in \{2, 3\}$.

The argument here is exactly the same, replacing $\tilde{\pi} = \frac{\pi(4)-1/2}{\phi}$ with $\breve{\pi} = \frac{\pi(4)-1/2-\gamma\frac{\varepsilon}{2}}{\phi(1-\varepsilon)} \approx \tilde{\pi}$, $\Delta(l_A, l_B)$ with $\Delta^F(l_A, l_B)$. Crucially, for every $\varepsilon > 0$, we still have

$$\lim_{\theta \to 0} \Delta^F(l_A, l_B) = \Delta^0(l_A, l_B).$$

With reference to the lists l', l'' and l''' invoked in the proof of 2, we therefore have that (owing again to the LK voters):

$$\lim_{\theta \to 0} \Pi_J(l''', l_{-J}) - \Pi_J(l'', l_{-J}) > 0 \to \lim_{\theta \to 0} \Pi_J^F(l''', l_{-J}) - \Pi_J^F(l'', l_{-J}) > 0$$

and

$$\lim_{\theta \to 0} \Pi_J^F(l^*, l_{-J}) - \Pi_J^F(l', l_{-J}) \geq \lim_{\theta \to 0} \Pi_J(l^*, l_{-J}) - \Pi_J(l', l_{-J}).$$

Part (ii)

(A) Re-express the difference

$$\Pi_A(l^*, l^*) - \Pi_A(l', l^*) - (\Pi_B(l^*, l^*) - \Pi_B(l', l^*))$$

as

$$
\begin{aligned}
&[p_A(4|l^*,l^*) - p_A(1|l^*,l^*)][\tilde{G}(4|l^*,l^*) - \tilde{G}(4|l',l^*) - (\tilde{G}(1|l^*,l^*) - \tilde{G}(1|l',l^*))] \\
&+ [p_A(3|l^*,l^*) - p_A(2|l^*,l^*)][\tilde{G}(3|l^*,l^*) - \tilde{G}(3|l',l^*) - (\tilde{G}(2|l^*,l^*) - \tilde{G}(2|l',l^*))] \\
&+ \sum_{k=1}^{4}[R_A^\theta(k;l^*,l',l^*) - R_B^\theta(k;l^*,l',l^*)]\tilde{G}(k;l',l^*).
\end{aligned}
$$

The argument is unchanged, and we obtain

$$
\begin{aligned}
&\Pi_A^F(l^*,l^*) - \Pi_A^F(l',l^*) - (\Pi_B^F(l^*,l^*) - \Pi_B^F(l',l^*)) \\
&= [p_A^F(4|l^*,l^*) - p_A^F(1|l^*,l^*)][\tilde{G}(4|l^*,l^*) - \tilde{G}(4|l',l^*) - (\tilde{G}(1|l^*,l^*) - \tilde{G}(1|l',l^*))] \\
&+ [p_A^F(3|l^*,l^*) - p_A^F(2|l^*,l^*)][\tilde{G}(3|l^*,l^*) - \tilde{G}(3|l',l^*) - (\tilde{G}(2|l^*,l^*) - \tilde{G}(2|l',l^*))] \\
&+ \sum_{k=1}^{4}[R_A^{F\theta}(k;l^*,l',l^*) - R_B^{F\theta}(k;l^*,l',l^*)]\tilde{G}(k;l',l^*).
\end{aligned}
$$

(B) Show that as θ approaches zero, $[p_A(4|l^*,l^*) - p_A(1|l^*,l^*)]$ and $[p_A(3|l^*,l^*) - p_A(2|l^*,l^*)]$ approach zero. Again, since

$$\lim_{\theta\to 0}\Delta^F(l^*,l^*) = \lim_{\theta\to 0}\Delta^* = 0$$

we still obtain that $[p_A^F(4|l^*,l^*) - p_A^F(1|l^*,l^*)] \underset{\theta\to 0}{\longrightarrow} 0$ and $[p_A^F(3|l^*,l^*) - p_A^F(2|l^*,l^*)] \underset{\theta\to 0}{\longrightarrow} 0$.

(C) Show that when $-\tilde{\pi} - \Delta^* \leq \overline{\xi}$ (Case 1) or when $-\tilde{\pi} - \Delta^* > \overline{\xi}$ (Case 2), $p_A(3|l^*,l^*) - p_A(2|l^*,l^*)$ and $p_A(4|l^*,l^*) - p_A(1|l^*,l^*)$ are strictly positive.

The argument extends verbatim in both cases, replacing $\tilde{\pi} = \frac{\pi(4)-1/2}{\phi}$ and Δ^* with $\breve{\pi} = \frac{\pi(4)-1/2-\gamma\frac{\varepsilon}{2}}{\phi(1-\varepsilon)} \approx \tilde{\pi}$ and $\Delta^F(l^*,l^*) \approx \Delta^*$.

(D) Show that for $k \in \{3,4\}$, $p_A(k|l^*,l^*) - p_A(5-k|l^*,l^*)$ converges to zero more slowly than

$$R_A^\theta(k;l^*,l',l^*) - R_B^\theta(k;l^*,l',l^*)$$

and

$$R_A^\theta(5-k;l^*,l',l^*) - R_B^\theta(5-k;l^*,l',l^*).$$

so that as θ approaches zero, $\Pi_A(l^*,l^*) - \Pi_A(l',l^*) - (\Pi_B(l^*,l^*) - \Pi_B(l',l^*))$ approaches zero from above.

The argument extends verbatim, replacing R_A^θ, R_B^θ, $\tilde{\pi}(k)$, Δ'_A, Δ'_B, and Δ^* with $R_A^{F\theta}$, $R_B^{F\theta}$,

$\breve{\pi}(k) \approx \tilde{\pi}(k)$, $\Delta^F(l', l^*) \approx \Delta'_A$, $\Delta^F(l^*, l') \approx \Delta'_B$, and $\Delta^F(l^*, l^*) \approx \Delta^*$. In particular, notice that

$$\frac{d\tau_k(\Delta, \gamma, \theta)}{d\theta} = \frac{d}{d\theta} \left\{ \frac{\psi(1-\theta) + \theta \tilde{f}\left(\breve{\pi}(k, \gamma) - \Delta\right)}{\sum_{r=2}^{4} \psi(1-\theta) + \theta \tilde{f}\left(\breve{\pi}(r, \gamma) - \Delta\right)} \right\} = \frac{2\tilde{f}(\breve{\pi}(k)) - \sum_{r \neq k} \tilde{f}(\breve{\pi}(r))}{9\psi}$$

which allows us to claim that $\lim_{\theta \to 0} \frac{\Delta^F(l', l^*)}{\Delta^F(l^*, l^*)} < \infty$ and $\lim_{\theta \to 0} \frac{\Delta^F(l^*, l')}{\Delta^F(l^*, l^*)} < \infty$. Moreover,

$$\lim_{\theta \to 0} \Delta^F(l^*, l') = \lim_{\theta \to 0} \Delta^F(l', l^*) = \lim_{\theta \to 0} \Delta^F(l^*, l^*).$$

Hence, the rest of the argument follows the same steps of the proof of Proposition 2. □

Appendix D: Additional Empirical Results, Sensitivity Tests and Robustness

(i.) The Marginal Rank Hypothesis: Closed versus Flexible Lists. Supplemental Appendix C formally establishes that our key results (Propositions 1 and 2) should hold in both closed and flexible list contexts. Nonetheless, Figure S.3 splits our sample into elections that were held under closed lists (1991 and 1994) and elections held under flexible lists (1998 to 2014). We observe identical nomination patterns both before and after the introduction of preference votes. However, our estimates are less precise in the earlier period, in which we have both fewer elections but also less coverage of the military draft data.

One subtle difference between our predictions under closed and flexible lists is that under *closed* lists, the marginal rank hypothesis further predicts that a party's best candidate should be placed in a contested rank; in Supplemental Appendix C, we show that under *flexible* lists the marginal rank hypothesis is consistent with a party's best candidate being placed in the top rank (i.e., *Capolista*). That is, the hypothesis predicts that: under *closed* lists: $\max\{\hat{\beta}_1, \hat{\beta}_2\} < \min\{\hat{\beta}_3, \hat{\beta}_4, 0\}$, while under *flexible* lists: $\max\{\hat{\beta}_1, \hat{\beta}_2\} > \max\{\hat{\beta}_3, \hat{\beta}_4, 0\}$. This makes closed lists an especially salient subset of the data for testing the implications of the marginal rank hypothesis. Nonetheless, inspection of Figure S.3 reveals that average candidate quality is maximized in the top two rank categories under *both* pre- and post-1998 electoral periods. This further encourages us to reject the marginal rank hypothesis.

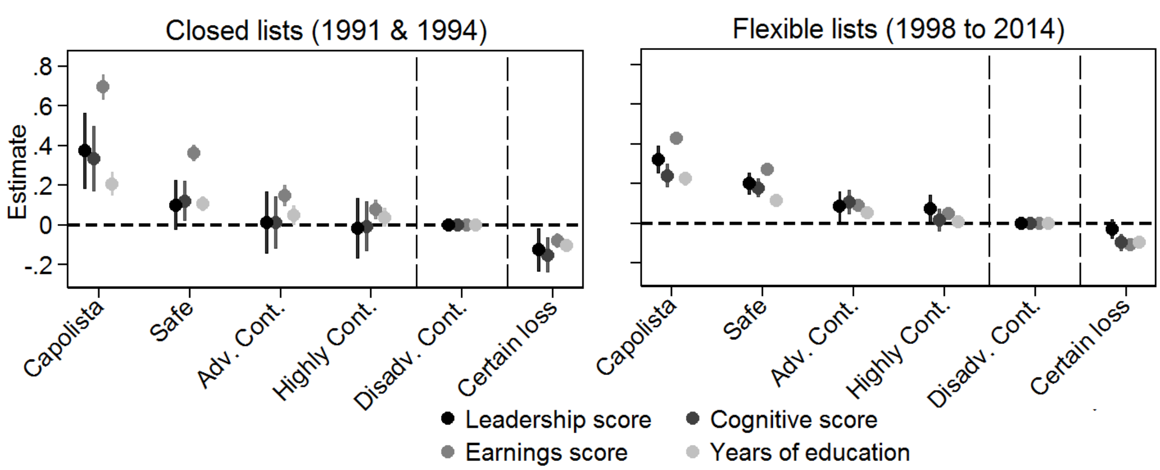

Figure S.3 – Average valence levels across ballot rank categories with *closed* and *flexible* lists. The figure replicates the analysis in Figure 4 in two sub-samples of data, split according to whether elections where held with closed or flexible lists.

(ii.) Alternative Measures of Advantaged and Disadvantaged Parties. Our benchmark presentation defines advantaged versus disadvantaged parties by reference to their share of seats

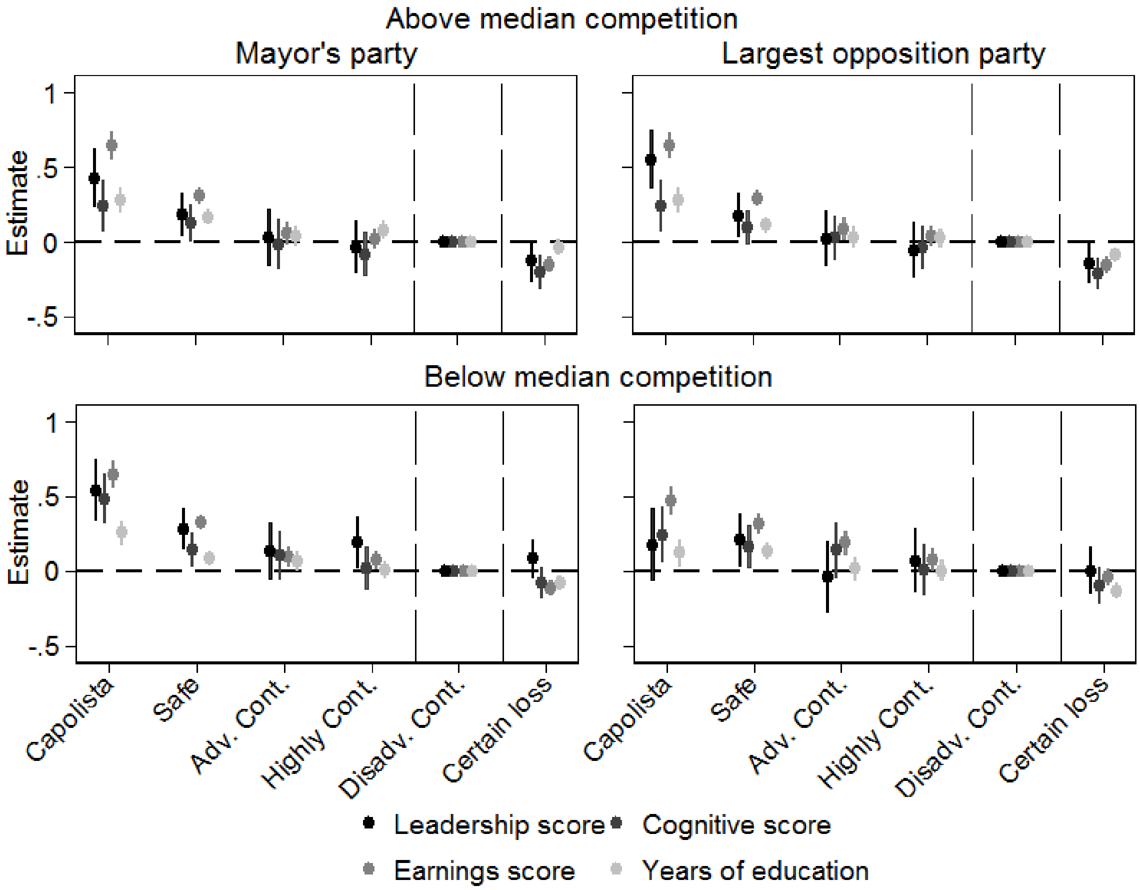

Figure S.4 – **Average valence levels across ballot rank categories**. Average valence levels across ballot rank categories in high and low political competition, and in governing and opposition parties. The figure replicates the analysis in Figure 5 in the 2*2 division of sub-samples. The top two plots show the estimates in municipalities with above-median political competition, and the bottom two plots for below-median political competition. The two plots on the left include the party of the mayor, and the two plots to the right includes the largest opposition party.

above or below the median seat share in our sample, i.e., in Figure 5.

To show that our main result i.e., that the tendency to place better candidates in higher ranks is most pronounced in *advantaged* parties, we consider an alternative sample split, by dividing districts into *above median* and *below median* competition. We define the extent of competition in a municipality as the average vote share difference between the two party blocs over the previous three elections: a small vote share difference reflects higher competition between the blocs, since the distribution of votes is more even across them. We also focus on two classes of political parties: those that appointed the mayor in the previous electoral cycle, and the largest opposition party from the previous electoral cycle, i.e., the largest party that was not part of the governing coalition. We reestimate equation (7) using this alternative sample split in Figure S.4.

While all parties appear to place better candidates in safer ranks, the weakest relationship holds for opposition parties in low competition settings. These constitute the parties that are least likely to appoint the chief executive, and thus the figure is consistent with the Top-Down Rank Order Hypothesis, highlighting the robustness of our empirical findings to this alternative sample split, and their consistency with our theoretical predictions.

(iii.) What if Advantaged Parties Have Access to Better Candidates? Recall that Propositions 1 and 2 are derived in a theoretical framework in each party has access to candidates of the same quality. In practice, however,electorally advantaged parties may have access to a better supply of candidates than electorally disadvantaged parties. Even if both parties assign better candidates to better ranks, we might still observe a stronger rank versus average quality gradient amongst advantaged parties simply on the basis of different pools of candidates.

To see this point, suppose that we were to compare two parties: one that we identified as *advantaged* and one that we identified as *disadvantaged*. Suppose that an advantaged party A has access to a candidate pool $Q_A = \{1, 2, 3, 4, 8\}$ and disadvantaged party B has access to $Q_B = \{1, 2, 3, 4, 5\}$. Suppose further that both parties assign candidates to ranks in order of quality, i.e., $l_A(1) = 8$, $l_A(2) = 4$, and so on, while $l_B(1) = 5$, $l_B(2) = 4$, and so on. Recall that our baseline empirical specification is:

$$Y_i = \alpha_{p,t} + \beta_1 \mathit{Capo}_i + \beta_2 \mathit{Safe}_i + \beta_3 \mathit{Adv\ cont}_i + \beta_4 \mathit{Highly\ cont}_i + \beta_5 \mathit{Certain\ loss}_i + \varepsilon_i. \tag{S.24}$$

Even if both parties have precisely the same propensity to assign better candidates to higher ballot ranks, we would obtain a higher estimate $\hat{\beta}_1$ amongst advantaged parties simply because the average candidate quality in those ranks reflects a better supply of talent.

To explore whether this is true, we plot our raw mean quality measures in Figure S.5. With the exception of the *Earnings* score, we see no evidence that average candidate quality within a given ballot rank category is larger in the advantaged parties—in fact, we observe the opposite! To see why this could arise even if the parties have homogeneous talent pools, consider a hypothetical ten-seat district in which we categorize the disadvantaged party's ranks two and three as "safe", but categorize the advantaged party's ranks two through 5 as "safe". Even if both parties have access to the same set of candidates with quality $1, ..., 10$, and both parties assign better candidates to better ranks, the average quality of candidates in safe ranks would be $\frac{9+8}{2} = 8.5$ in the below median size party, and would be $\frac{9+8+7+6}{4} = 7.5$ in the above median size party.

For all measures other than Years of Education, we observe that the valence score of the highest-ranked candidate—the "*Capolista*"—is significantly lower in uncompetitive opposi-

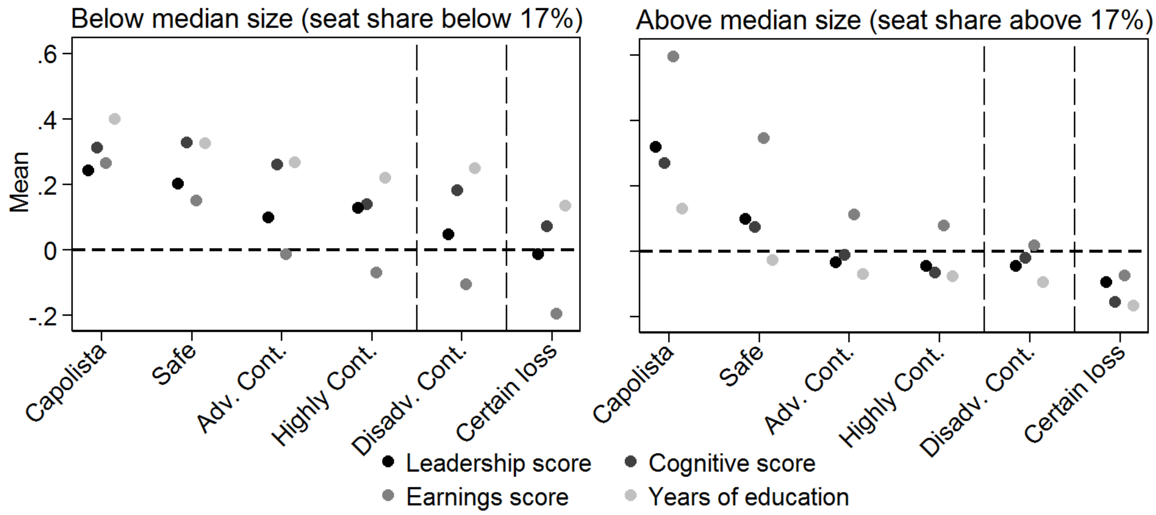

Figure S.5 – Mean valence levels across ballot rank categories. The analysis replicates the analysis in Figure 5 but uses shows the average of the four valence measures on the y-axis instead of the regression estimates.

tion parties. Nonetheless, in the remaining ballot categories we do not see a clear discrepancy across the four categories.

Nonetheless, we propose an alternative test that can accommodate differences in average candidate quality across different ballot ranks of different parties. For each ballot rank $k \in \{1, ..., n-1\}$, on a party's electoral ballot, we compute the share of ranks $k+1, ..., n$ whose candidates' valence scores are less than or equal to the valence score of the candidate in rank k. For each rank k, this produces a "dyadic rank score" between zero and one; higher values imply fewer violations of the *top-down rank order* principle that better candidates should be bound in higher ranks on the party's ballot. Formally, letting $q(k)$ denote the quality of a candidate in rank k, the *dyadic rank score* of rank $k \in \{1, ..., n-1\}$ on a party's ballot l is:

$$s_l(k) = \frac{\sum_{k'=k+1}^{n} \mathbf{1}[q(k) \geq q(k')]}{n-k}. \tag{S.25}$$

To understand the usefulness of this measure, observe that it depends only on the *ordering* of candidate quality within a party's ballot, not the quality *level*. For example, the top-down rank-order hypothesis predicts that $s_l(k) = 1$ for every rank k and every ballot l. In the example above, this is precisely the dyadic rank score that we estimate, despite the fact that the parties' quality pools differ.

Figure S.6 plots estimated coefficients from equation (7) using dyadic rank scores, i.e., it shows the average difference in dyadic rank scores between each ballot category relative to

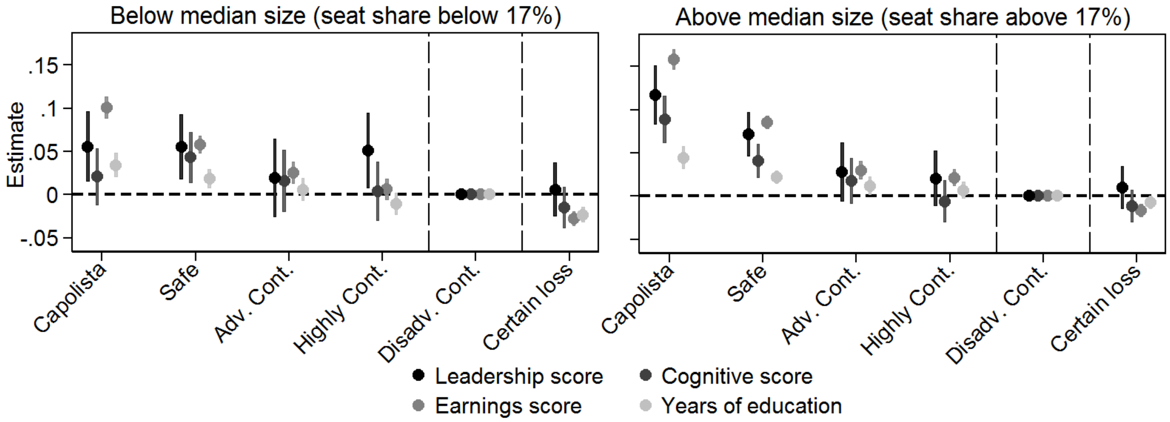

Figure S.6 – Dyadic valence across ballot rank categories. The figure shows estimated coefficients from Equation (7), where the dyadic valence difference between each category of ballot ranks is estimated relative to the reference category of the disadvantaged competitive ranks. The outcome variables are dyadic scores for each the four valence measurements. Vertical lines show 95% confidence intervals. All regressions include fixed effects for every local party in every election period. The data is all candidates in local parties with at least one candidate in each of the six ballot rank categories. The number of observations in the regression analysis is: Leadership Score: $N = 35,662$; Cognitive Score: $N = 48,088$, Earnings Score: $N = 209,995$, Years of Education: $N = 185,382$.

the reference category of disadvantaged competitive ranks. It therefore replicates Figure 5 in our benchmark empirical specification but using the average dyadic rank score in each ballot category, rather than average candidate quality in each ballot category. The figure highlights that our main results—falsifying the marginal rank hypothesis and providing support for both the main and ancillary predictions of the top-down rank order hypothesis—extend to this alternative measure.

(iv.) Using Alternative Information to Predict Election Probabilities. We show that our results are not sensitive to the information that we use to predict the election probabilities for each list rank. Recall that our benchmark empirical model uses data from the previous two elections to forecast a party's prospects in next election. We use two alternative sets of data for this analysis, which we present in Figure S.9. First, we only use data from the previous election, presenting the results for this analysis on the left-hand side of Figure S.9. Second, we remove the interaction terms between previous seats won and council size from the prediction regression and rely solely on previous seats won, presenting the results for this analysis on the right-hand side of Figure S.9. This analysis provides more or less identical results as our main analysis. This highlights that our findings are not sensitive to the choice of data that is used to construct electoral forecasts.

(v.) Different Cut-off Thresholds for Defining Rank Categories. We return to using the seat

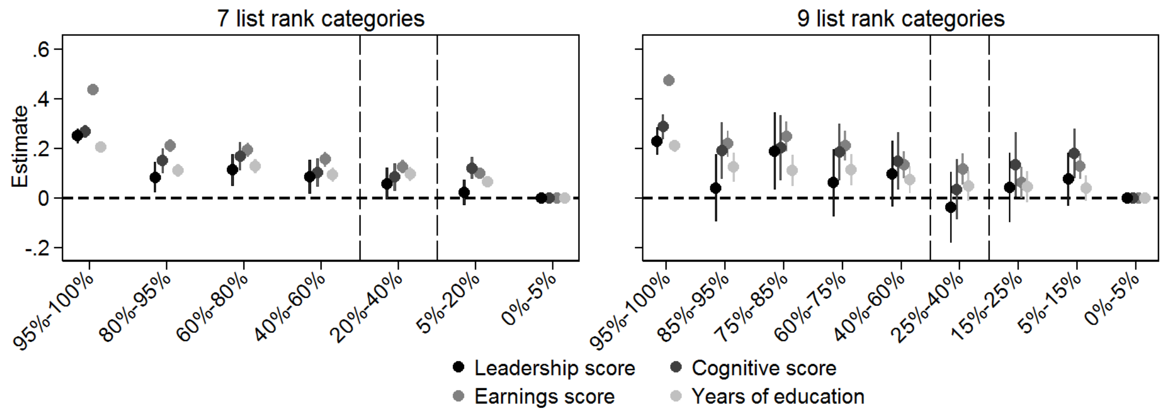

Figure S.7 – **Average valence levels across ballot rank categories, using alternative rank category classifications.** The figure replicates the analysis in Figure 4, but uses different criteria for assigning ranks to different categories. The analysis in the left-hand figure uses 7 categories and the right-hand side 9 categories.

prediction model from our benchmark model but instead change the probability cut-offs that are used to define the various list rank categories. We consider two alternative definitions in this analysis. When making these division we have to consider that the cumulative distribution is steeper around 50%. To obtain a reasonably sized sample in each probability category we therefore use wider intervals around 50%. Also, to ensure sure that we always have an observation in our reference category we change the reference category to those with a prospect of election at or below 5%—which contains the most candidates.

We show the results of this analysis in the Figure S.7. In the left-hand figure, we divide the sample into 7 categories and in the right-hand side we divide the sample into 9 categories. We find no support for the marginal rank hypothesis. We noted that the estimates are noisier, for which there are two reasons. First, there are fewer individuals in each category. Secondly, there are fewer lists for which we have an individual in each category.

We further highlight that our results are not driven by how we define the list categories by constructing a new variable intended to capture a candidate's relative list rank, which is simply list rank minus the seats won by the party in the last election. For each relative rank, we then define a dummy variable. Within each party, we estimate the competence at each rank versus the reference category, which we define as relative rank zero. We restrict the sample to parties that won 4 or more seats in the last election and to ranks that we have in all party groups. With 4 seats as the size cut-off, this means that we can go to a relative rank of -3—for example, in groups with 4 seats this is the top-ranked candidate. Since we use a much smaller subset of individuals in each list, we do not include fixed effects for local party. Instead we redefine the outcome variable as the difference with the local party mean.

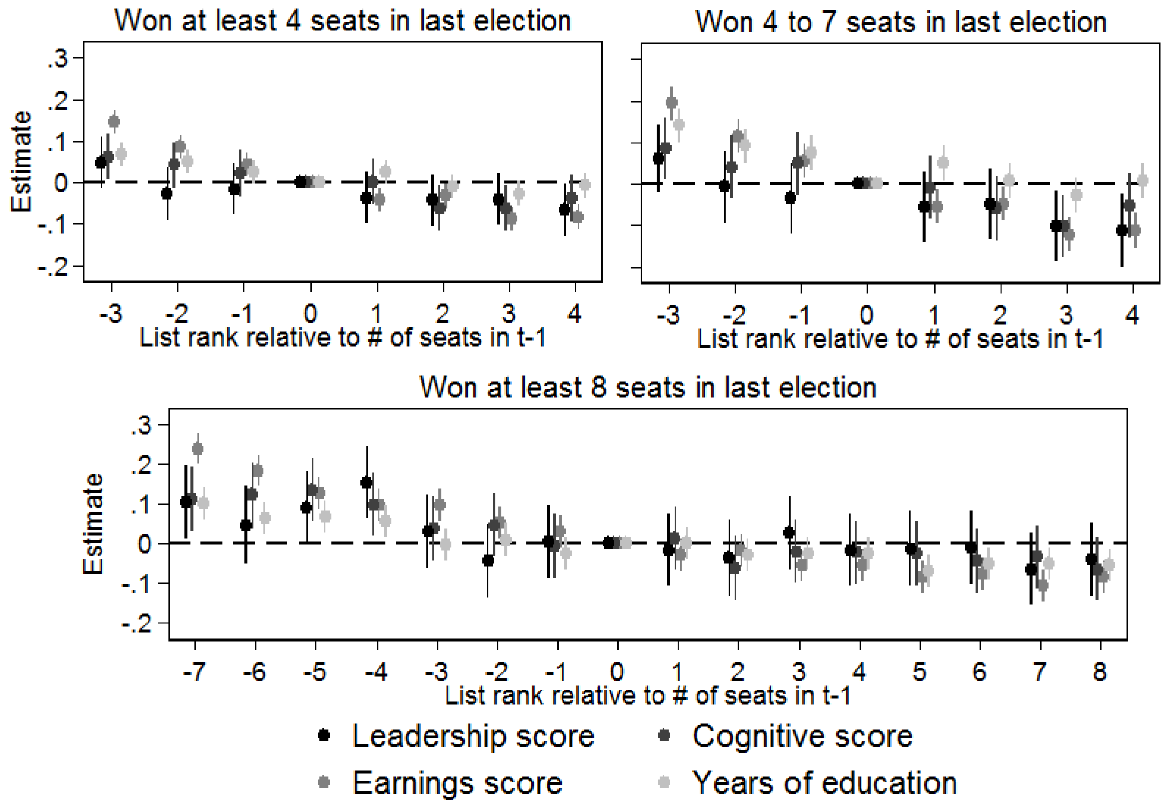

Figure S.8 – **Average valence levels based on relative list rank.** The figure shows estimated coefficients from Equation (7), but uses list rank relative to the number of seats won in the last election instead of list rank categories. The outcome variables are the four valence measurements, defined as the difference to the local party mean. Vertical lines show 95% confidence intervals. The data is all parties that won at least 4 seats in the last election (upper left figure) between 4 and 7 seats (upper right figure) and at least 8 seats (lower figure).

We present the results in Figure S.8. In the top left figure, we show the results for the full sample. In the top right figure, we restrict the sample to smaller party groups, i.e. those that won between 4 and 7 seats in the last election. In the bottom figure we use large parties, i.e those that won at least 8 seats in the last election. In the last sample we can expand the number of list ranks used to -7 (this is the relative rank of the capolista in parties that won 8 seats in the last election). Although the estimates are noisier than our main analysis, we still see a clear downward sloping trend for competence as we move done the list, thereby yielding little support for the marginal rank hypothesis.

(vi.) Different Measures of Candidate Quality. We use alternative measures of candidate quality in Figure S.10. In the left-hand side of the figure we use incumbency as a characteristic that may either be directly valued by voters, or may be correlated with underlying

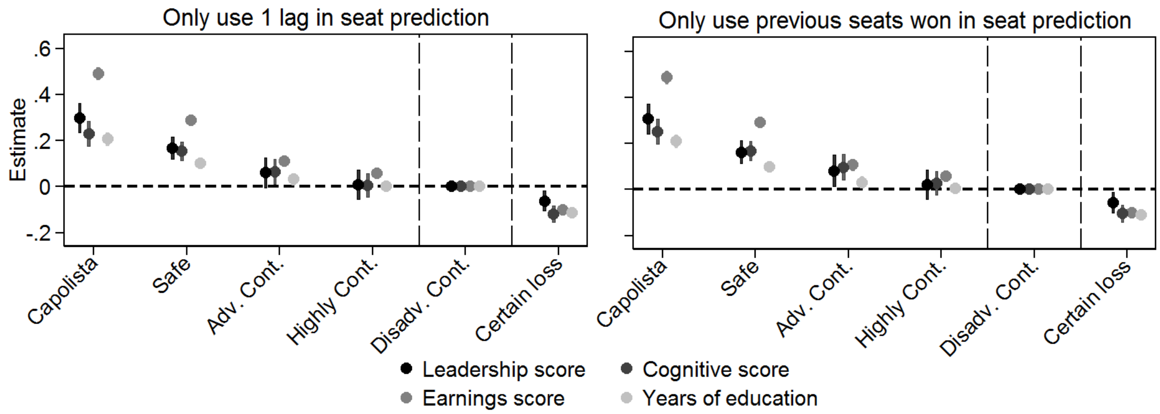

Figure S.9 – Average valence levels across ballot rank categories using alternative predictive data. The figure replicates the analysis in Figure 4, but uses different data to predict the probability of winning a seat. The analysis in the left-hand side figure relies on data from the most recent election and the right-hand side figure only relies on data on previous number of seats won.

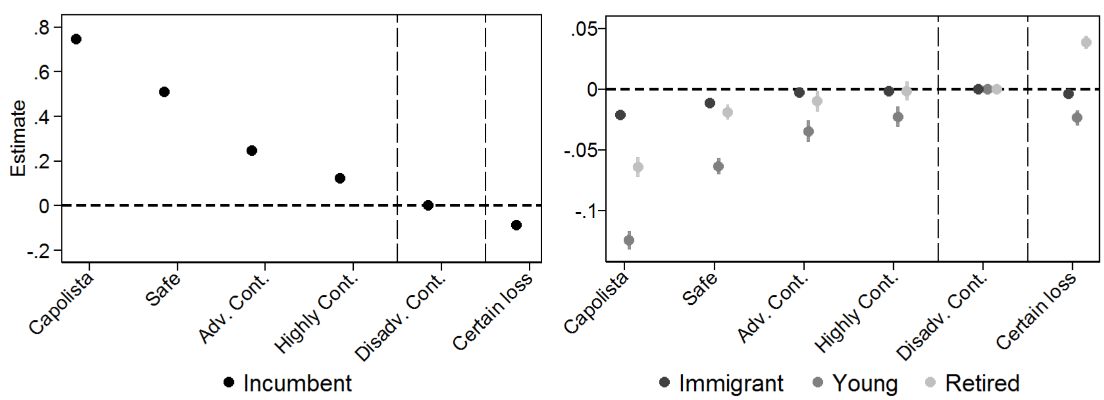

Figure S.10 – Candidate selection levels across ballot rank categories using alternative quality measures. The figure shows estimated coefficients from Equation 7, but uses different measures of candidate quality. The outcomes are incumbency (left hand side), being young (less than 35 years), above retirement age (65 years) or a non-nordic immigrant (the latter three on the right-hand side).

qualities . Incumbency has been used as a valence measure, for example, in Hirano and Snyder (2014), and is also a proxy of standing within the party. Neither for incumbency do we find and support for the marginal rank hypothesis as the share of incumbents is clearly, and strictly decreasing as we move down the list. The other three outcomes, being young (less than 35 years), above retirement age (65 years) or a non-nordic immigrant are also related to preference votes.[19] For none of these alternative measures do we see any non-monotonicity within the contested ranks.

[19] Conditional on list rank both young candidates and non-nordic immigrants receive more preference votes, while the retired receive less support. Results are available from authors upon request.

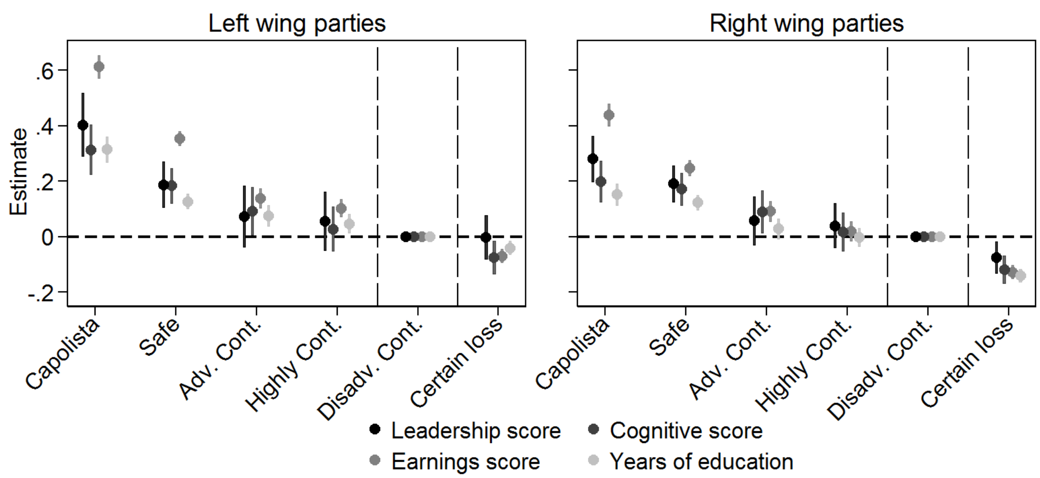

Figure S.11 – **Average valence levels across ballot rank categories in *Left wing* and *Right wing* parties.** The figure replicates the analysis in Figure 4 in two sub-samples of data, split by party ideology.

(vii.) Sample Split by Party Ideology. Since Swedish parties differ in their nomination procedures—for example, left-wing parties being more likely to employ internal primaries—we might expect the quality-rank gradient to differ across parties that use different nomination processes, regardless of their local electoral strength. In Figure S.11, we split our sample by ideology, and observe that center-left and center-right parties exhibit similar quality-rank gradients. Both figure are consistent with the Top-Down Rank Order Hypothesis, but inconsistent with the Marginal Rank Hypothesis.