Essays on Electoral Institutions

Leonardo Carella

Mansfield College

Thesis submitted in partial fulfilment of the requirements for the degree of

DPhil in Politics

in the Department of Politics and International Relations at the University of Oxford

Michaelmas Term 2022

Word Count: 54,163

Abstract

Do electoral systems matter for the outcomes of democratic politics beyond their well-known relationship to party systems? This thesis presents three studies on the consequences of electoral institutions, highlighting how they shape incentives and opportunities for political actors operating at the intra-party level. The first article considers how the spatial representativeness of legislatures – i.e. the extent to which MPs from different parts of the country are equally represented – varies with constituency and ballot structure. The paper proposes a novel measurement to gauge the descriptive representation of places in parliaments and develops a theoretical framework linking spatial representation to parties' incentives and voters' ability to elect local MPs. Consistently with the theory, the analysis finds that mixed-member systems and preferential voting mechanisms are associated with more geographically representative parliaments. The second paper asks whether sub-national legislators' likelihood to run for the national parliament is influenced by the electoral system through which they were elected. Leveraging within-legislature variation in electoral rules across German State parliaments, I find that list-PR MPs are more likely to attempt 'level hopping' than single-member district legislators. It is argued that this depends on the lower levels of electoral security of the former, and to the more direct accountability of the latter to local (s)electorates. The third paper develops a theoretical model of preference vote distribution in preferential-list PR system, extending the approach behind Taagepera and Shugart's Seat Product Model to intra-party competition. I show that the share of preference votes for the first-ranked and last-eligible candidates, as well as the effective number of candidates in a list, can be accurately predicted in expectation as functions of institutional variables. Overall, the thesis makes contributions in terms of measurement, data and theory towards a broader and richer understanding of the intra-party effects of electoral institutions.

Acknowledgements

As most projects started in the Autumn of 2019, this thesis had a bumpy ride. If it has come together in the end and I have made it through all in one piece, it in no small part due to the help, support and encouragement of many. I wish to thank in particular my supervisor Ben Ansell and co-author Andy Eggers for their advice, insights and patience. I could not have wished for better mentors. I owe them a debt I can never pay back, but only endeavour to pay forward.

My doctorate could not have been possible without the funding of the Economic and Social Research Council, and a generous maintainance grant from the Bank of Italy.

Various parts of the thesis benefitted from the comments, suggestions and criticism of Tarik Abou-Chadi, Spyros Kosmidis, Matthew Shugart, Tom O'Grady, Mats Ahrenshop, Marta Antonetti, Ashrakat Elshehawy and Jonne Kamphorst. I am grateful to them for lending me their time and brainpower. I also wish to acknowledge the helpful feedback from the participants to the Comparative Research on Political Elites Conference (Hertie School), the Inequality and Representation Workshop (Oxford), the Democracy, Elections and Comparative Politics PGR Conference (Manchester), and the anonymous reviewers at the *British Journal of Political Science*.

I am grateful to the many people who aided in the process of data collection, including Till Heinsohn, Miroslav Nemčok, Christian Lessmann, Erzen Öncel, Patricia Calca, Sebastian Koehler, Jaehee Jun, Lewis Baston, Esme Lillywhite and a regiment of employees of German Federal and State Election Commissions. Special thanks go to Mats Ahrenshop and Ashrakat Elshehawy for their translation work.

I am thankful to Ezequiel Gonzalez-Ocantos, David Rueda, Jane Gingrich, Jody LaPorte, David Doyle, Scot Peterson and Rob Ford for giving me opportunities for professional growth through my doctorate; and to my students, for reminding me why I want to grow into this profession in the first place.

To my friends Ash, César, Christine, Francesco, Geoff, Giuseppe, Jamie, Luis, Marta, Mats, Nick, Nicole and Tom – thank you. For the emotional support, for the stimulating conversations, for the coffee breaks, for the late-night takeaways, for the banter and for the gossip. You made a world of difference.

Finally, I wish to thank my parents Daria and Donato, who made huge sacrifices to let me continue my adventure abroad, put up with me during lockdown and endured my thesis crises. None of this would be possible without your love and confidence in me. Through it all, you have always been on my mind.

Questa tesi è dedicata a mamma e papà.

Contents

List of Abbreviations

AfD	Alternative for Germany
c	number of candidates in a list
c^*	number of pertinent vote-earning candidates
CDF	cumulative distribution function
CDU	Christian Democratic Union
CLPR	closed-list proportional representation
CSU	Christian Social Union
d	deviation from prediction index
EMD	Earth Mover's Distance
FDP	Free Democratic Party
FLPR	flexible-list proportional representation
GLP	Global Leadership Project
HYDE 3.2	History Dataset of the Global Environment
M	district magnitude
MP	member of parliament
MTM	multi-member
MXM	mixed-member

N_c . . . . . effective number of candidates

N_S . . . . . effective number of parties

OLPR . . . . . open-list proportional representation

p . . . . . number of preference votes

PDS . . . . . Party of Democratic Socialism

PLPR . . . . . preferential-list proportional representation

PR . . . . . proportional representation

PV . . . . . preferential voting

r . . . . . over-nomination ratio

S . . . . . assembly size

s . . . . . number of seats attributed to a list

σ_1 . . . . . fractional share of seats of the largest party

SBW . . . . . Shugart-Bergman-Watt

SM . . . . . single-member

SMD . . . . . single-member district

SNTV . . . . . single non-transferable vote

SPD . . . . . Social Democratic Party of Germany

SPM . . . . . seat-product model

STV . . . . . single transferable vote

SURLI . . . . . Spatial Un-Representativeness of Legislatures Index

v_1 . . . . . share of preference votes of the first-ranked candidate

v_n . . . . . share of preference votes of the n^{th} ranked candidate

v_s . . . . . share of preference votes of the last eligible candidate

Chapter 1

Introduction

What are the consequences of electoral institutions for the outcomes of democratic politics? Insofar as elections determine who gets represented in a polity and who governs it, almost any upshot of democratic politics can be said to be related – more or less directly – to the relatively small set of 'rules of the game' under which these contests take place. Indeed, studies aimed at answering the many variants of the question of electoral system effects constitute one of the richest and most fruitful lines of research in contemporary political science. In this sense, this thesis builds on sturdy foundations. There is a broad consensus, for instance, about how district magnitude can be expected to shape key features of a party system: number of parties, party sizes, and disproportionality between votes and seats (Amorim Neto and Cox, 1997; Powell Jr and Vanberg, 2000; Shugart and Taagepera, 2017). We also have a reasonable understanding of how these proximate effects of electoral institutions translate into more 'downstream' political outcomes, such as parliament-executive relationships (Hallerberg, 2004; Lijphart, 2012), redistributive politics (Iversen and Soskice, 2006; Persson and Tabellini, 2005) and turnout rates (Blais and Carty, 1990; Brockington, 2004; Eggers, 2015).

The aim of the three papers that this thesis comprises is not so much to challenge this influential body of knowledge; rather, they aim to expand the comparative electoral system research agenda to new or relatively understudied dependent variables. Specifically, I show how electoral institutions can be theoretically and empirically linked to three outcomes: (1) the descriptive representation of places in parliaments, (2) the progressively ambitious behaviour of sub-national legislators, and (3) the extent of intra-party competition in preferential-list systems. In all three cases, the theories tested centre *intra-party* mechanisms through which "variations in electoral rules affect the internal organisation of parties and the ways in which individual

legislators (or candidates) relate to constituents" (Shugart, 2005, p. 36). This stands in contrast with the more widely studied *inter-party* mechanisms, which refer to "how electoral systems affect the translation of votes into seats for competing political parties, and [...] the overall nature of the party system" (Shugart, 2005, pp. 29-30). Concretely, the intra-party mechanisms emphasised in the body of the thesis concern how electoral rules exert an influence on the selection of candidates according to their socio-demographic characteristics, the behavioural incentives of individual representatives once elected, and competition between co-partisans within a party list. Therefore, the common contribution of these papers lies in highlighting that electoral systems affect political outcomes not just by determining the number of parties and degree of proportionality between them, but also by changing the strategic environment in which aspirants to political office, electoral candidates, elected representatives and party selectorates operate.

A second common thread running through the three studies is the central role of political actors' *expectations* in determining how variation in electoral rules translates into systematic differences in incentives and behaviour. The argument that "expectations play a central role in coordinating electoral activity and choice" (Cox, 1997, p. 8) is as old as the study of electoral systems itself (most notably in Duverger, 1959, but a version is already in Droop, 1881). In developing this insight further, this thesis shows just how pervasive the effect of expectations set by institutional environments can be, engendering mechanisms that reach into the workings of a political system beyond the well-trodden path of strategic voting and strategic party entry. For instance, parties' expectations over the marginal gains from selecting a certain candidate over another, which are conditional on electoral rules, will shape the demographic composition of legislative assemblies down the line. Furthermore, legislators' expectations over their own electoral security (and which gate-keepers matter for it) shape career choices and elite circulation in predictable ways. Finally, the expectations that candidates in a preferential list system have over how their list will fare on the inter-party dimension affects the perceived viability of their candidacy, the extent to which they will vie for votes, and thus the overall competitiveness of the list on the intra-party dimension.

The first paper, coauthored with Andrew Eggers and presented in chapter 3, asks which electoral institutions favour the selection of a political class that is representative of a polity's geographic diversity. There is a normative case for investigating the descriptive representation of geographical extraction in parliaments – just as a growing body of scholarship has done for ethnic groups, gender and class – as extant research shows that voters prize highly 'localness' as a trait in candidates, and that politicians' ties to locales are consequential for policy outcomes. The theoretical argument developed posits that equal representation of members of

parliaments from different parts of the country should be a function of (1) the incentives parties have of fielding candidates representative of their districts and (2) the leverage voters have to overrule parties' selection when they 'parachute' a non-local candidate in favour of a local one. Drawing on municipality-of-birth data for over 13,000 legislators in 62 democratic legislatures, the paper proposes a Spatial Un-Representativeness of Legislatures Index (SURLI), a measurement which solves a number of potential issues of cross-country comparability related to variation in countries' geographical shape and size.

It is shown that mixed-member constituency structures – and, to a lesser extent, the availability of preferential voting mechanisms – are associated with more spatially representative legislatures. This is consistent with our theoretical priors, as mixed-member constituency structures reduce two possible sources of geographical bias: low visibility of candidates and low competitiveness of district elections. Within-case analysis on British and German single-member district legislators further explores the mechanisms behind mixed-member systems' overperformance relative to single-member plurality. The finding that non-competitive district elections are less likely to return a local-born member of parliament (MP) in 'pure' single-member districts but not in the single-member tier of a mixed-member system is consistent with the theory. The results have normative implications for electoral system choice, insofar as they challenge the 'constituency linkage' argument for majoritarian systems like single-member district plurality: as far as the descriptive dimension of representation is concerned, these institutions do not seem to guarantee better representation of geographical locales in legislatures.

The second paper, in chapter 4, asks whether legislators elected under single-member district plurality and list proportional representation (PR) behave differently with respects to their 'progressive ambition': that is, their ambition to move up to a higher territorial level of politics. To answer this question, the paper examines candidacies of sitting German State legislators to the Federal Parliament across ten elections (1987-2021): I define these candidacies as 'level-hopping attempts'. The German context allows to leverage within-legislature variation in the nature of legislators' electoral connection to their district due to the widespread adoption of mixed-member systems in sub-national legislatures. There are theoretical reasons to hypothesise that State MPs elected from party lists have stronger incentives to attempt level-hopping than MPs elected in the single-member districts tier. First, list PR MPs have lower re-election prospects at State level, as they cannot rely on the incumbency advantage to the same extent as single-member district legislators and are unlikely to have a secure alternative pathway to re-election in the single-member district tier. This decreases the prospective value of staying in the current post relative to moving up to the Federal parliament (the

'electoral security' mechanism). Secondly, single-member district MPs have stronger incentives to cater to their local constituency: a candidacy for a higher position while in office comes with reputational costs that may endanger their future in the State parliament if such candidacy is unsuccessful (the 'constituency commitment' mechanism).

The empirical analysis proceeds in two steps. First, I test the assumption that list PR legislators face lower re-election rates, and find strong evidence in this sense. Then, I present a series of regression models where the likelihood of observing a level-hopping attempt is predicted as a function of legislators' tier of election and other individual-level covariates, and repeat this exercise distinguishing between 'secure' and 'insecure' candidacies. I find that list PR legislators are indeed more likely than single-member district MPs to display progressively ambitious behaviour. Moreover, the difference in propensity to attempt level-hopping between legislators from the two tiers is substantively larger and more significant for insecure candidacies than secure ones. This is also consistent with the argument that signalling the intention to leave a legislative office for a more prestigious one is particularly costly *if unsuccessful*, and that the expected costs are relatively higher for district MPs. The main implication of the results is that patterns of upwards mobility of legislative elites in multi-level polities are influenced by sub-national electoral institutions, insofar as these determine re-election prospects of legislators and the nature of their mandate accountability in future periods. The analysis also contributes to the debate on the 'mandate divide' in mixed-member systems, highlighting the linkage between electoral tiers and an aspect of legislators' behavioural repertoire – their propensity to attempt level-hopping – that has not been previously identified in the literature.

The third paper, in chapter 5, asks how the distribution of preference votes in a preferential-list proportional representation (PLPR) system *should* vary according to electoral institutions. The levels of analysis – the electoral list and the electoral district – thus strike a middle course between the 'macro' level of the first paper, where the unit of observation is a parliamentary assembly, and the second one, which individuates 'micro'-level mechanism affecting the conduct of individual legislators. Moreover, in contrast to both, it makes first and foremost a theoretical contribution, elaborating at length on the question of how things *should be* – in this case, what intra-party competition should look like – rather than simply aiming to find out how they are. Developing theory-building techniques employed by authors such as Rein Taagepera and Matthew Shugart in their work on the effects of electoral institutions on inter-party competition, the third paper presents a series of *quantitatively predictive logical models* of quantities of intra-party competition. This approach does not simply aim to derive directional hypotheses as to the expected effect of a

configuration of electoral system quantities on an outcome, but rather to specify a functional form of these relationships. Concretely, in the paper I propose that (1) the share of preference votes obtained by the top candidate in a seat-winning list, (2) the effective number of candidates in a seat-winning list, and (3) the share of preference votes obtained by the last-eligible candidate in a seat-winning list can be predicted in expectation as functions of the product of quantifiable institutional variables. These are: (a) the district magnitude, (b) the number of preference votes available to voters and (c) a parameter expressing the legal constraint on over-nominations as the ratio between the maximum number of candidates a party can field and the district magnitude.

The quantitative predictions are tested on preference vote data from 31 elections across nine countries, for a total of over 2,600 seat-winning lists accounting for the intra-party performance of over 57,000 candidates. The first empirical section shows that the functional forms derived describe accurately the overall relationship between the product of institutional variables and the target characteristics of intra-party competition. Moreover, the predictions are substantially less biased than those of an existing model of intra-party competition (Shugart, Bergman and Watt, 2013) and perform roughly equally well across open- and flexible-list systems. The second part of the empirical section compares the intra-party models for the preference share of the top-ranked candidate and the effective number of candidates in a list with their inter-party counterparts: the seat-product model predictions for the seat share of the largest party and the effective number of parties in an assembly (Taagepera, 2007; Shugart and Taagepera, 2017). It is shown that the precision of the intra-party models in predicting district-level median values of the quantities of interest is comparable to that of the more well-established inter-party analogues. The paper concludes by illustrating how the functional forms derived may be used to inform institutional choice, by making the case to broaden the study of intra-party competition to encompass the full diversity and complexity of PLPR sub-types, and by identifying possible avenues to improve further the models.

Chapter 2

Literature Review

While the importance of electoral institutions in shaping patterns of democratic competition was already a major theme in the work of party systems scholars of the 1950s and 1960s (Duverger, 1959; Rae, 1967; Sartori, 1968), it is no understatement to say that the field has been revolutionised towards the end of the twentieth century.[1] In the space of twenty years, comparative electoral systems research progressed from "the most underdeveloped subject in political science" (Lijphart, 1985, p. 3) to a "mature field" (Shugart, 2005, p. 25). This wave of scholarship (Lijphart, 1984; Shugart and Taagepera, 1989; Cox, 1997; Norris, 2004) accomplished two remarkable achievement for the political science as a whole. First, it has been able to derive theoretically a set of 'interlocked' equations that describe empirical regularities in the relationships between features of the electoral systems and political outcomes such as effective number of parties and disproportionality (Shugart and Taagepera, 1989; Taagepera, 2007; Shugart and Taagepera, 2017). These predictive statements, more akin to the laws of the natural sciences than to common forms of directional theorising in the social sciences, represent a major advancement of political science as a *scientific* endeavour. Secondly, and perhaps even more remarkably, electoral system research can boast to have become "one of the portions of comparative politics that has been most consumed by real-world practitioners" (Shugart, 2005, p. 27). By drawing connections between formal rules and *desiderata* of a democratic systems, this body of knowledge has been influential in shaping constitutional design in newly democratised polities and institutional reform of established democracies (Carey et al., 2013).

As it would be impossible to do justice to all the debates in this vast literature, the following sections focus

[1] It is probably no coincidence that the surge of interest in electoral system effects ran parallel to the historical phase of institution-building that followed the 'third wave' of democratisation.

on three areas of research that are closely related to the analysis presented in this thesis, highlighting how each of the three papers contributes to debates in these subfields. These correspond to works linking electoral institutions to (1) descriptive representation, (2) legislators' behaviour, and (3) intra-party competition.

2.1 Descriptive Representation

The notion of descriptive representation was most notably articulated by Hanna Pitkin in terms of a correspondence between the ascriptive traits of the representatives and the represented:

> It depends on the representative's characteristics, on what he *is* or is like, on being something rather than doing something. The representative does not act for others; he "stands for" them, by virtue of a correspondence or connection between them, a resemblance or connection. In political terms, what seems important is less what the legislature does than how it is composed. (Pitkin, 1967, p. 61)

Though for Pitkin this connection is not sufficient for democratic representation, advocates of the 'politics of presence' (Mansbridge, 1999; Lovenduski and Norris, 2003) have drawn on her conceptualisation to argue that socio-demographic imbalances in the composition of elites are products of injustice to be addressed for a more effective aggregation of interest in democratic politics. Empirical literature has indeed shown that the degree to which political elites reflect their electorates in terms of e.g. gender, race, and class has important consequences for policy outcomes (Duflo, 2012; Carnes and Lupu, 2015) and voter-level political behaviour (Dassonneville and McAllister, 2018; Heath, 2018).

The bulk of the literature on electoral institutions and descriptive representation has focussed on gender, often starting from the observation that women's presence in elective bodies tends to be higher in high-district magnitude proportional representation than under systems with single-member or low-magnitude districts (McAllister and Studlar, 2002). This relationship seems to be primarily driven by intra-party effects on candidate supply, though the literature is still far from a consensus on the relative importance of at least four distinct hypothesised mechanisms (Schwindt-Bayer, 2005; Dhima et al., 2021). First, multi-member districts with sufficiently large district magnitudes allow parties to maxisimise their social appeal by diversifying their candidate slate, as opposed to single- or small-magnitude districts, where parties can aim to elect at most one candidate and therefore tend to have a bias towards male insiders (Norris, 2006, p. 41). As Marsh (1988, p. 280) put it, "qualities which selectors might feel, accurately or otherwise, could be electoral

liabilities in a party's sole candidate, like being a woman or a member of an ethnic minority, are needed for purposes of balance when several are being picked." Secondly, Matland and Studlar (1996) show that the lower political cost of ticket-balancing in systems with high district magnitude sets in motion a 'contagion' process, whereby mainstream parties respond to minor parties' promotion of women representation by increasing female candidate recruitment. Thirdly, just as PR systems allow parties to balance tickets informally due to centralised candidate selection, these institutional environments also lend themselves to the introduction of formal gender quotas and placement mandates (Schmidt, 2009). Finally, proportional systems have higher rates of legislator turnover: by favouring new entrants and undermining incumbents, these systems accelerate the translation of changes in social norms around female politicians into shifts in political personnel (Schwindt-Bayer, 2005). The relationship between ballot structure within non-majoritarian systems and women's descriptive representation is fundamentally contested: the 'old' consensus (Rule, 1994; Wängnerud, 2009) in favour of open-list proportional representation (OLPR) has come increasingly under scrutiny by a new wave of studies that find a null or negative relationship between preferential voting and women's representation (Htun, 2005; Bieber and Wingerter, 2020; Gonzalez-Eiras and Sanz, 2021), or a more complex one where the direction of the effect is conditional on contextual social norms (Valdini, 2012; see also Moser and Scheiner, 2012, pp. 208-235, for an analogous argument about mixed-member system tiers).

There is also a smaller and more fragmentary literature on electoral institutions' effects on minority ethnic representation. While cross-country generalisations are obviously difficult due to the diversity of polities' ethnic compositions, these works provide some evidence that electoral institutions affect minorities' descriptive representation both via inter-party and intra-party mechanisms. Non-permissive electoral systems on the inter-party dimension – e.g. single-member district (SMD) plurality, or systems with high legal thresholds – may constrain the descriptive representation of minorities, preventing the formation and success of ethnic-interest parties (Friedman, 2005; Lublin, 2017). On the intra-party dimension, the core claim in the literature is that single-member districts favour the descriptive representation of *large and geographically concentrated* groups: "the election of ethnic minorities in SMD systems tends to be based on geographic concentration and the ability of a minority group to constitute a critical mass within a given electoral district" (Moser, 2008, p. 289; see also Zollinger and Bochsler, 2012). While there is evidence that ethnic concentration provides incentives for parties to place minority candidates in favourable positions in list-based systems as well (Geese and Schacht, 2019), studies on mixed-member systems tend to confirm that spatially clustered ethnic minorities are elected in the single-member tier at higher rates than in the proportional

tier (Kostadinova, 2007; Moser, 2008). Just as in the case of gender, there is no consensus on whether preferential voting mechanisms are beneficial to ethnic minority representation: any effect of these institutions is likely highly idiosyncratic to the nature of a country's ethnic heterogeneity and intergroup relations. Some authors argue that OLPR or flexible-list proportional representation (FLPR) systems are similar to single-member district plurality insofar as they make electoral contests 'candidate-centred' rather the 'party-centred', thus incentivising parties to field minority candidates to benefit from their ability to mobilise co-ethnics (Dancygier, 2017; Janssen, 2022). Others individuate potential 'ethnic penalties' at the ballot box as a potential downside of preferential voting (Portmann and Stojanović, 2019).

The first paper of this thesis contributes to this literature draws on this literature to analyse empirically how electoral systems are related to the descriptive representation of legislators' geographical extraction. Childs and Cowley (2011) have made a compelling argument for the desirability of descriptive representation of 'the local', arguing that many of the normative arguments advanced in favour of gender and ethnicity also apply to geography: "in an era of identity politics, where voters see themselves as having a local identity, it should come as no surprise that localness is raised as an identity that warrants descriptive representation" (Childs and Cowley, 2011, p. 14). The argument presented in the paper starts from the assumption that parties' and voters' preferences over legislators' localness are, to some extent, in conflict. Parties' candidate selection is biased towards certain parts of the country due to the unequal spatial distribution of resources and credentials that drive political ambition, while voters are biased towards candidates from their part of the country. To the extent that electoral institutions solve this tension in favour of voters, we should observe geographically representative legislative elites. The analysis in section 3.4.3 suggests that mixed-member (MXM) constituency structures and, to a lesser extent, preferential voting (PV) are effective in overcoming the two main hurdles to spatially balanced local representation: low visibility of candidates in multi-member (MTM) districts and high levels of seat safety in single-member (SM) districts. This conclusion has normative implications, insofar as it casts doubt on the argument that majoritarian electoral systems ensure tighter linkages between representatives and territorial units within a polity (Blais, 1991).

2.2 Legislators' Behaviour

Alongside electoral systems' influence on the socio-demographic characteristics of representatives, a second major strand in the literature on intra-party effects focusses on how electoral institutions affect their behaviour once they take up office. To the extent that legislators can be assumed to seek re-election as one of

their primary aims (Mayhew, 1974), their priors over the circumstances under which they may achieve such goal will be forcefully shaped by the rules under which future electoral contests are fought. In particular, the scholarship has identified two main mechanism through which electoral institutions affect legislators' behaviour in between elections: they shape patterns of *accountability* that MPs are subject to, and the *competitiveness* of the electoral contests ahead of them (André, Depauw and Shugart, 2014). The deriving incentive and opportunity structures contribute to resolve tensions between legislators' own preferences, those of their party and those of voters, often interacting with other institutional factors – e.g. candidate selection rules and parliamentary procedure (Carey, 2007; Itzkovitch-Malka and Hazan, 2017; Shomer, 2017). Broadly speaking, the main theoretical contention in the literature is that institutional environments that increase the personal accountability of a legislator to their district's voters and intra-party competition for seats create incentives for legislators to adopt 'personal vote-seeking' behavioural repertoires (Carey and Shugart, 1995; Cain, Ferejohn and Fiorina, 2013), while those that empower party leaders in their ability to reward and sanction legislators should favour party discipline.

The most widely studied aspect of legislators' behaviour is what is commonly thought to be their primary function: parliamentary voting. Studies leveraging within-legislature variation in electoral rules suggest a relationship between 'candidate-centred' systems – STV, SNTV, OLPR and, though to a lesser extent, FLPR and SMD plurality – and independent-minded behaviour of legislators, in contrast with stricter party unity under 'party-centred' systems such as closed-list proportional representation (CLPR) or plurality-at-large. For instance, Hix (2004) finds that defection from party group vote in the European Parliament is higher for MEPs elected under PLPR than for those elected under CLPR. Moreover, Däubler and Hix (2018) employ the case study of the Christian-Democratic European Parliament grouping's vote on a gay rights resolution to show that, even across different types of preferential-list PR system rules, seat-allocation rules that empower voters over parties are associated with higher responsiveness to their constituents' opinion. Findings from mixed-member systems, where theory suggests lower rates of defection in the (party-centred) closed-list PR tier than in the (relatively more) candidate-centred single-member district tier, are however more ambiguous. While there is some evidence for this 'mandate divide' hypothesis (Ohmura, 2014; Batto, 2012), it is by no means consistent across contexts (Kerevel, 2010; Jun and Hix, 2010; Rich, 2014), and perhaps the framework is of limited relevance outside mixed-member systems with very low levels of party system institutionalisation (Thames, 2005).

Other empirical work suggests caution in overstating the role of electoral institutions as major determi-

nants of legislative behaviour. After all, legislators may operate under the assumption that voters' attention to roll-call voting is simply too low to justify breaking the party line, especially as this comes with immediate non-electoral costs for the individual MP (e.g. being passed over for promotion) and for the party (e.g. loss of agenda control) (André, Depauw and Shugart, 2014, pp. 242-243). As Martin (2014, p. 477) put it, "what goes on within the legislature may be more important for influencing legislators' behaviour than what goes on at the ballot box". The work of Scott Morgenstern on Latin American parliaments, for instance, finds only limited evidence in favour of a consistent association between electoral institutions and levels of party unity at cross-country level, which instead is shown to be more forcefully driven by ideological and electoral characteristics of parties as well as the form of the executive (Morgenstern, 2003; Morgenstern and Swindle, 2005). Studies that leverage electoral system change also suggest little change in patterns of voting behaviour in parliaments when new electoral rules are introduced (Coman, 2012; Olivella and Tavits, 2014).

Nonetheless, legislators seem to be more responsive to electoral system incentives for personal-vote seeking on other aspects of their behaviour, which may be more effective and less costly signals of their constituency orientation. Most notably, electoral institutions seem to matter for *constituency service*, a label encompassing all "activities addressing the nonpolicy grievances or looking out for the nonpolicy interests of citizens in the member's district" (Crisp and Simoneau, 2018, p. 345). For instance, findings from PARTIREF (André and Depauw, 2013; André, Freire and Papp, 2014), a large-scale cross-national legislator survey, confirm systematic differences in behaviour between legislators elected under party- and candidate-centred systems, as well as providing evidence in favour of an interactive effect of district magnitude and candidate-centred nature of the ballot formula (as hypothesised by Carey and Shugart, 1995). The time a legislator spends in their district, their efforts to promote local interests and the intensity of their interaction with local organisations decrease with district magnitude in closed-list systems and increase in open list systems. These insights into legislators' behaviour chime with PARTIREF's findings on their *attitudes*, which suggest that MPs in candidate-centred systems perceive their constituents as the focus of their mandate, as opposed to representatives in party-centred systems, who see tend to see themselves as making the interest of 'all people' or that of their own party (Dudzińska et al., 2014). Alongside constituency service, committee allocation also seems to be systematically related to electoral institutions. Evidence from mixed-member systems suggests that, where legislators elected under different sets of electoral incentives serve in the same parliament, they select committees in ways that are theoretically consistent with the notion that electoral rules shape legislators' perception of their representative *foci*. Single-member district MPs tend to sit in

bodies that allocate resources that can be targeted to districts, while those elected under closed-list PR rules tend to work on areas that are core to their party's electoral appeal (Stratmann and Baur, 2002; Maaser and Stratmann, 2018).[2]

The contribution of this thesis's second paper to this literature consists in analysing an overlooked aspect of legislators' behaviour: their propensity to run for 'higher' office in the middle of a legislative term. In the context of a multi-level system, I define this phenomenon as a 'level-hopping attempt'. While there is some research on the relationship between electoral institutions and career trajectories (Jones et al., 2002; Høyland, Hobolt and Hix, 2019), this is to my knowledge the first study that analyses 'progressively ambitious' (Schlesinger, 1966) behaviour in mixed-member legislatures at sub-national level. Theoretically, level-hopping attempts are distinct from the other aspects of legislative behaviour reviewed thus far in two senses. First, the main tension in legislators' incentive structure is not between interests of constituents and of the party, but between constituents' preferences and the legislator's own career ambitions. Secondly, the mechanism through which electoral institutions influence representatives' choices is not just one related to their accountability, but has also (and perhaps mainly) to do with the competitiveness of their re-election contests. In other words, sitting legislators who try to 'jump ship' midway through a parliamentary term may not do so only because they believe that they are shielded from voter accountability, but also because they might fear that they will not be able to retain their seat in future elections. The analysis of candidacies of German State legislators for the Federal parliament in section 4.7 suggest that closed-list PR incentivises level-hopping attempts relative to single-member districts, but the effect is smaller for Federal candidacies in secure positions.

2.3 Intra-Party Competition

Intra-party competition – the extent to which elections represent contests between co-partisans, taking place in parallel to the more visible contest between parties – has mostly been studied as a mediating rather than an outcome variable. A variety of theoretical and empirical work on electoral system effects in fact assume intra-party competition as a *mechanism* linking electoral institutions and political outcomes. For instance, electoral institutions may engender different incentives and opportunities for candidates to differentiate themselves from co-partisans in parliamentary elections. In this perspective, it has been argued that systems enabling intra-party competition tend to shift patterns of electoral choice from partisan to valence

[2]For a more detailed review of the literature on legislators' behaviour in mixed-member systems, see section 4.2.

considerations (Riera and Cantú, 2022), increase within-party positional heterogeneity (Catalinac, 2018), hamper incumbency advantage (Ariga, 2015; Passarelli, 2020), as well as reducing partisan coordination in voter mobilisation and turnout (Robbins, 2010; Söderlund, 2017). Moreover, competition between co-partisans present parties with personnel management and collective action problems. As coordination issues tend to be greater for larger parties, it has been variously claimed that permissive systems on the intra-party dimension present favourable conditions for small parties. Under candidate-centred non-pooling systems like single non-transferable vote (SNTV), larger parties have more room for over- and under-nomination error, resulting in a small-party bias relative to PR in the attribution of seats relative to pooling systems (Shugart and Taagepera, 1989, 170; Pachón and Shugart, 2010 though see Cox, 1996 for a partial critique). A similar issue arises in open-list PR systems with coalitional lists, where smaller parties within a list are more adept at optimising preference vote distribution than larger ones (Calvo, Guarnieri and Limongi, 2015). Finally, permissive electoral rules on the intra-party dimension have been linked to the institutionalisation of competition between co-partisans in the form of highly factionalised parties (Grofman et al., 1999; Boucek, 2009; Invernizzi, 2021).

In terms of the institutional determinants of intra-party competition, the seminal contribution is an the influential paper by Carey and Shugart (1995), which proposes a rank ordering of electoral formulas according to the extent to which they induce candidates to seek personal votes as opposed to advancing the party's collective reputation. Their conceptualisation of intra-party competition in terms of 'incentives to cultivate a personal vote' has set the tone for much of the subsequent debate, although – as noted by Grofman (2005) – the notion conflates politicians' incentives to cater to a local constituency and incentives to develop a reputation distinct from the party brand. In the paper, Carey and Shugart (1995) argue that personal vote-seeking behaviour depends on three variables: (1) party leaders' control over candidacies and ballot ranking (higher $\rightarrow$ more party-centred), (2) whether votes for individual candidates are pooled by list (party-centred) or not (candidate-centred), and (3) whether voters express a preference only for an individual (most candidate-centred), for both a list and an individual, or only for a list (most party-centred). Moreover, they famously conjectured that in candidate-centred systems incentives to cultivate a personal vote increase with district magnitude, while in party-centred systems they are decreasing in the same variable. This hypothesis has subsequently been subjected to a number of theoretical qualifications and revisions (Crisp, Jensen and Shomer, 2007; Kselman, 2020; Buisseret et al., 2022), as well as a great deal of empirical testing on various proxies for and consequences of personal vote-seeking, with some notable theory-confirming results

(Shugart, Valdini and Suominen, 2005; Chang and Golden, 2007; André, Depauw and Shugart, 2014).

The third paper of this thesis takes a less well-trodden path in the study of intra-party competition: instead of identifying variation in personal vote-seeking behaviour, it aims to describe and predict the *distribution* of the personal vote across candidates as a function of party and institutional variables. To do so, I propose a model of intra-party outcomes of elections in preferential-list PR (PLPR) systems, a highly diverse and increasingly common (Renwick and Pilet, 2016; Cheibub and Nalepa, 2020; Passarelli, 2020) electoral system family "in which inter-party allocation takes place across party lists, but voters are permitted (or sometimes required [...]) to indicate a preference for one or more candidates within one list" (Shugart, 2005, p. 40). Two of the quantities that can be predicted with this model – the preference vote share of the first-ranked candidate and the effective number of candidates – are effectively indicators of intra-party competitiveness in expectation. The starting point of the paper is the work of Shugart, Bergman and Watt (2013), who draw on the theoretical toolkit behind the seat-product model of inter-party competition (Taagepera, 2007; Shugart and Taagepera, 2017) to model first- and last-elected candidate shares in open-list PR (a subtype of PLPR) and SNTV (a non-list preferential voting system). Their model is further developed to reflect the fact that the 'co-partisan crowdedness' of a list, expressed by the number of candidates fielded, is not the only structural driver of intra-party competition: as argued by Crisp, Jensen and Shomer (2007), the expected number of seats at stake for a list is just as important to incentivise personal vote-seeking. By justifying theoretically quantitative benchmarks of intra-party competition quantities and testing them on a diverse sample of elections, the paper makes a number of contributions to the literature. First, the model specifies the relationship between district magnitude and intra-party competition as one proceeding via two avenues: the proliferation of candidates *and* the increase in the expected number of seats attributed to lists as magnitude increases. Secondly, on the basis of the empirical findings I can conclude that – once we account for the number of preferences available to voters – preference vote distribution follows a roughly similar logic in both main subtypes of PLPR: open- and flexible-list PR. Finally, it is shown that the model can be used to select preference thresholds in flexible-list systems, so as to obtain a desired balance-of-power between parties and voters in determining intra-party electoral outcomes.

Chapter 3

Electoral Systems and Geographic Representation: The Descriptive Representation of Places in Parliaments

Electoral Systems and Geographic Representation:

The Descriptive Representation of Places in Parliaments

Leonardo Carella[†]

Andrew C. Eggers[‡]

Abstract

Who gets represented in legislatures, and how does this depend on electoral institutions? Others have asked this question from the perspective of gender, race, and class; we focus on *space*, asking whether MPs disproportionately come from some places rather than others and how this depends on electoral rules. Using data on over 13,000 legislators in 62 democracies, we develop a new measure of the extent to which the spatial distribution of MP birthplaces matches the spatial distribution of the citizens they represent. Contrary to received wisdom, we find that single-member district systems do not have more geographically representative parliaments than multi-member district systems, while mixed-member systems perform significantly better than both. We attribute the higher spatial representativeness of mixed-member systems to contamination effects in their single-member tier, and we present evidence for this explanation from within-country analysis of the UK and Germany.

[†]Mansfield College and University of Oxford. email: leonardo.carella@mansfield.ox.ac.uk
[‡]University of Chicago. email: aeggers@uchicago.edu

3.1 Introduction

One of the key insights of the literature on political elites is that compositional differences between the demographic makeup of representatives and that of their voters have important consequences for the quality of democratic representation (Bratton and Ray, 2002; Preuhs, 2005; Carnes, 2012). While comparative scholarship in this field has made substantial advances in mapping representational gaps along the dimensions of gender, race and class origins, this paper aims to investigate *to what extent legislatures around the world reflect the geographic diversity of the voters they represent.* We do so by developing and computing a comparable measure of inequalities in the descriptive representation of *places* in parliaments, the Spatial Un-Representativeness of Legislatures Index (SURLI). Alongside this descriptive exercise, we theorise and investigate how electoral systems may explain cross-country variation on this variable, drawing on the common finding in the descriptive representation literature that different electoral institutions produce distinct opportunities for social groups to access political offices.

Our theoretical starting point is the idea that voters and party elites have divergent preferences when it comes to the geographical ties of elected officials. Voters tend to prefer candidates with local ties, all else equal, perhaps because they assume that a local candidate is more likely to share their preferences (Shugart, Valdini and Suominen, 2005). For example, a recent conjoint experiment in the UK showed that localness mattered to voters 15 times more than biological sex (Campbell and Cowley, 2014, p. 758); Cowley (2013) also reports that for British voters having a representative 'from the same area' as them comes a close second to 'someone with the same political viewpoint', and way ahead of a representative sharing respondents' class, sex, religious or racial characteristics. Party elites who select candidates, on the other hand, place a lower weight on localness relative to other traits such as social connections, political experience, expertise, or education; in some cases, fielding a local candidate requires sacrificing on these other attributes. How does this tension between voters' and parties' preferences get resolved? Parties seeking to win seats may field local candidates to cater to voter demands, and voters may use the ballot to select local candidates over non-local ones, but the extent to which either occurs likely depends on the electoral system. Below we offer a theoretical lens through which the relationship between electoral systems and the geographical representativeness of legislatures can be better understood. We also investigate this relationship focusing on two aspects of variation in electoral systems: constituency structure and ballot structure (Reeve and Ware, 2013). We offer further evidence of a connection between electoral systems and geographical representativeness using both a cross-country analysis and within-country analysis in Britain and Germany.

To briefly preview our findings, we observe that (perhaps surprisingly) single-member (SM) district systems do not have more geographically representative legislatures than multi-member (MTM) distict systems. Although the large number of districts in SM district systems is supposed to ensure a representative legislature if parties fielded local candidates in each district, parties may not have much incentive to do so, especially in a safe district. We do find, however, that mixed-member (MXM) systems tend to have more geographically representative legislatures than either of those systems; we suggest this may reflect the incentives created by spillovers between tiers, which tend to heighten the incentives for parties to field a local candidate, and increase opportunities for voters to select one, in the nominal (SM) tier. Our within-country analysis of constituency MPs in Britain and Germany provides evidence consistent with this hypothesis.

One of our main contributions is a new, cross-nationally comparable measure of legislatures' geographical representativeness, which we call the Spatial Un-Representativeness of Legislatures Index (SURLI). The basis of SURLI is the discrepancy between the distribution of legislators' birthplaces and the distribution of citizens' residences, measured using Earth Mover's Distance (EMD) (Lupu, Selios and Warner, 2017). Heuristically, this captures the minimum amount of travel necessary to send an equal number of citizens to each MP birthplace. To make SURLI comparable across countries, we normalise each country's score by comparing it to the distribution of scores we obtain by repeatedly drawing random (and perfectly representative in expectation) hypothetical parliaments. Thus we obtain a measure that captures the degree of spatial bias in each country's legislature net of differences in the size and shape of the territory.

The paper is structured as follows. In section 3.2 we locate our work in the context of existing scholarly works on political geography, electoral behaviour and descriptive representation. In section 3.3 we discuss theoretically the possible channels through which features of electoral systems might influence how places are represented in a legislature. Section 3.4 tests these expectations on a cross-country sample and presents this paper's main empirical contribution: the measurement of SURLI for 62 legislatures, and the results of a cross-country regression that describes how it varies across electoral systems. In Section 3.5 we carry out within-country analyses on datasets of British and German constituency legislators, to assess a hypothesis about why single-member districts are more likely to produce local MPs in mixed systems than in single-member systems. Section 3.6 concludes by discussing briefly the normative implications of the study, its limitations, and further avenues for this research agenda.

3.2 Related Literature

Early democratic theorists worried about the 'problem of space' (Minicucci, 2001): how large representative democracies could aggregate disparate interests of communities located far apart, avoiding risks of secession or domination of one subunit over the other. Indeed, the territorial segmentation of the electorate in constituencies was often justified explicitly on the grounds that it would improve the quality of representation via the *localness* of candidates (Rehfeld, 2005). For instance, Montesquieu in *The Spirit of Laws* (1748) highlights the advantages of local deliberation for selecting representatives:

> One knows the needs of one's own town better than those of other towns, and one judges the ability of one's neighbors better than that of one's other compatriots. Therefore, members of the legislative body must not be drawn from the body of the nation at large; it is proper for the inhabitants of each principal town to choose a representative from it. (De Montesquieu, 1989, p. 159)

In a similar vein, Alexander Hamilton in *Federalist 36* (1788) touches on how representatives' local ties and knowledge enhance the responsiveness of the centre to the peripheries' needs, thus improving democratic outputs:

> If any question is depending in a State legislature respecting one of the counties, which demands a knowledge of local details, how is it acquired? No doubt from the information of the members of the county. Cannot the like knowledge be obtained in the national legislature from the representatives of each State? (Hamilton, Madison and Jay, 2008, p. 169)

Even as national parties have emerged and cleavages based on class and religion have become more salient, the notions of *democratic* and *territorial* representation have remained tightly linked in the electoral institutions of modern states. Non-geographical ways of dividing voters into distinct constituencies (such as the class franchises in 19th-century Prussia and Austria or Zimbabwe's 'white rolls') have always been rare and are almost unheard of in contemporary democracies.[1] Apart from a few countries including Israel and the Netherlands, geographically-disjoint electoral districts remain basic building blocks in every national electoral system. Thus we can still say, with Rehfeld (2005, p. 3), that "in almost every democracy in the world, citizens are represented by where they live."

[1] Hong Kong's 'functional constituencies' are a notable exception.

Correspondingly, *localness* continues to be an important prism through which voters evaluate candidates, as well as an electoral resource that candidates can count on in places where they have built long-standing personal and political networks. The electoral bonus candidates receive in 'their own beat' was famously described by Key (1949, p. 38) in *Southern Politics*:

> A candidate for governor normally carries his own county by a huge majority, and the harshest criticism that can be made of a politician is that he cannot win in his own beat or precinct. If his friends and neighbors who know him do not support him, why should those without this advantage trust a candidate?

There is substantial scholarly agreement that candidates receive an electoral boost in places where they have personal ties. Evidence in this direction abounds, from spatial ecological studies (Rice and Macht, 1987; Garand, 1988; Gimpel et al., 2008), experimental studies (Campbell and Cowley, 2014; Roy and Alcantara, 2015; Panagopoulos, Leighley and Hamel, 2017), and voter-level survey data (Johnson and Rosenblatt, 2006; Arzheimer and Evans, 2012, 2014; Evans et al., 2017). Moreover, the finding replicates in a variety of contexts: from the original testing grounds of the hypothesis in the United States (Rice and Macht, 1987; Garand, 1988; Gimpel et al., 2008) and Ireland (Gallagher, 1980; Marsh, 2007; Górecki and Marsh, 2012) to polities as diverse as Britain (Arzheimer and Evans, 2012, 2014; Evans et al., 2017), Japan (Horiuchi, Smith and Yamamoto, 2018), Canada (Roy and Alcantara, 2015; Blais and Daoust, 2017), Estonia (Tavits, 2010) and Norway (Fiva, Halse and Smith, 2018).

In contemporary scholarship, this effect is generally disaggregated into a *behavioural* and a *perceptual* component (Evans et al., 2017). The behavioural component refers to a greater ability of candidates to mobilise supporters in their immediate social networks, including those who live in the immediate surroundings of the candidate's place of residence (Górecki and Marsh, 2012). The perceptual component describes how localness serves as a heuristic for candidate desirability in voters' considerations: in low-information environments, someone from 'around here' can be more easily assumed to have the community's interests at heart than a 'parachuter' or a 'carpetbagger' (Campbell and Cowley, 2014; Campbell et al., 2019). In this sense, the micro-foundational mechanism is a form of in-group bias (Panagopoulos, Leighley and Hamel, 2017, pp. 867-868). Candidates are aware of this, and go to great lengths to cue their local credentials to voters. For instance, in a particularly striking example from the 2017 UK General Election, the Green Party candidate in Brent Central distributed campaign literature stating "I am a life long Brent resident, conceived in Harlesden, born in Kilburn, grew up in Queens Park and now reside Willesden" (Milazzo and

Townsley, 2018, p. 10).

Do politicians with local ties to a certain locale make for "better" representatives for that area? The evidence is mixed. Carozzi and Repetto (2016) find evidence from Italy (1994-2004) of pork-barrel spending driven by legislators born *outside* of their district in favor of their municipality of birth; Jennes and Persyn (2015, p 189) find that between 1995 and 2010 "per capita cash transfers to a Belgian electoral district are significantly higher for every federal minister originating from that electoral district". But there are also null findings in the literature. For instance, Fiva, Halse and Smith (2018) leverage close elections in Norway between 1953 and 2013, and find that representatives do *not* increase the level of investment in their hometown; Sällberg and Hansen (2020) find that localness was unrelated to the number of constituency mentions in the UK House of Commons over the 2015/2016 constituency session. Hence – as with other aspects of the 'politics of presence' (Preuhs, 2005) – the literature suggests that descriptive representation does not automatically translate into policy representation, but rather the relationship can be contextual and contingent on institutional factors.

Furthermore, this paper draws on comparative work on the relationship between electoral institutions and descriptive representation in legislatures. In this literature, proportional representation is often found to be conducive to demographically more balanced legislative assemblies. For instance, as far as the gender gap in political representation is concerned, "one of the most stable results in empirical research is that the election of women is favored by electoral systems with party lists, proportional representation (PR), and large district magnitudes" (Wängnerud, 2009, p. 54). PR systems seem to be also more inclusive of younger members of parliament (Joshi, 2013), as well as improving *policy* representation of low-income citizens (Carnes and Lupu, 2015; Bernauer, Giger and Rosset, 2015). Existing evidence on the effect of electoral systems on ethnic minority representation (Kostadinova, 2007; Moser, 2008; Wagner, 2014) is less clear-cut. Cross-national findings on electoral systems and the descriptive representation of *places* are scarce, not least due to the absence of comparable spatial measures of representational inequalities. Pedersen, Kjaer and Eliassen (2007) analysed *parachutage* – defined as the share of MPs residing outside the district at the time of election – in sixteen Western European countries, finding no clear association with electoral institutions. Perhaps the most influential work in this area is theoretical: Matthew Shugart and colleagues developed a series of models to derive candidates' incentive to seek a personal vote – and thus the likelihood of representatives being local – as a function of, among other factors, electoral rules. Their work suggests that the probability of a representative being local declines with district magnitude in closed-list PR, but

increases with district magnitude in open-list PR systems (Shugart, Valdini and Suominen, 2005; Carey and Shugart, 1995). André, Depauw and Shugart (2014) tested these hypotheses in three open-list PR countries (Finland, Luxembourg, Switzerland) and three closed-list PR (Spain, Portugal, Norway), and confirmed that the share of representatives who had held previous local office in the district varies as predicted.

3.3 Theory

How might electoral institutions affect geographical representation? Our theoretical starting point is that voters and parties have different preferences over the geographic extraction of legislators. As discussed in the previous section, there is strong evidence that voters prefer, *ceteris paribus*, legislators with strong ties to their local area. This local bias, however, contrasts with that of parties, which – in absence of electoral incentives to satisfy voters' preferences over candidates – may have incentives to field geographically biased candidates slates. This is for two reasons. First, parties are presented with a pool of aspirants that tends to over-represent certain parts of the country due to the unequal spatial distribution of wealth, influence, and social connections. Hence, even an entirely geography-blind process of candidate nomination may simply reproduce the bias in the supply of political ambition. Secondly, some candidate traits that parties tend to value – ideological profile, professional background, educational qualifications, employment within party structures, and ethnicity – are themselves likely to be distributed unevenly across the territory of a country. Hence, if parties can ignore voters' preferences, their ideal set of legislators will *at a minimum* reproduce the spatial inequalities of the pool of political aspirants; possibly, it may even aggravate them.

Whether the set of people actually elected to the legislature reflects such a territorial skew or the geographic balance preferred by voters depends, in our view, on how elections work. Electoral institutions are an important factor in determining how this tension is resolved: to the extent that they (1) induce viable seat-winning parties to field local candidates throughout a country (*party incentives*), and (2) allow voters to select locals over non-locals as their representatives (*voter leverage*), a legislature is likely to be geographically representative. In this section, we assess how these two criteria may vary depending on two dimensions of variation in electoral institutions:

— *Constituency structure* refers to the districting arrangements that define the geographical areas within which votes are translated into seats (Reeve and Ware, 2013; Shugart and Taagepera, 2017). The traditional distinction in this case is between single-member (SM) and multi-member (MTM) systems:

in the former, the votes in each constituency count towards the election of one representative; in the latter, they contribute to the election of more than one, up to the limit case where all MPs in the assembly are elected in the same nationwide district. A third type of constituency structure has become prominent since the 1990s: mixed-member (MXM) systems. These entail two spatially overlapping 'tiers' electing each a share of the assembly's legislators: one constituted by single-member (the nominal tier) and one by multi-member districts (the list tier).

— *Ballot structure* has been used to describe a range of options that voters have in making their choice in the polling booth. For the purposes of this analysis, we focus on a single key aspect: the presence or absence of *preferential voting* (PV), defined here in terms of any mechanism that allows voters to express a preference for a candidate that is meaningfully different from a preference for a party (Marsh, 1985; Farrell and McAllister, 2006). In single-member districts, parties field one candidate, so these never entail PV. Multi-member districts fall in the same category if they employ a closed-list system, whereby the party presents voters with an ordered list of candidates, and those in positions from 1 to n are elected, where n is the number of seats attributed to that party in the district. Conversely, under PV rules, voters have the option to have a say in the intra-party allocation of seats: either by voting for candidates *as well as* parties, and thereby changing the candidate ranking in a list system (e.g. Finland's open-list ballot), or by voting only exclusively for candidates in a non-list system (e.g. Ireland's single transferable vote ballot).

We begin by considering parties' incentives to field local candidates in a given district: fielding local candidates is neither necessary nor sufficient to achieve a nationally representative legislature,[2] but it is clearly an important ingredient. We assume, as noted, that at least in some situations parties would prefer to field non-local candidates; however, voters' preference for locals means that they could win over some voters by fielding local candidates instead. (We refer to voters who might be swayed by a candidate's localness as 'localist' voters.) Fielding a local candidate can thus be seen as similar to turnout mobilization or geographically targeted pork barrel spending: costly – because it entails sacrificing other candidate traits that a party *prima facie* values more than their geographical extraction – but potentially electorally rewarding. Electoral institutions matter, in this view, because they shape the electoral rewards parties stand to gain from fielding local candidates. In particular, assuming that parties want to win seats, the rewards of fielding

[2]With high malapportionment or geographically large districts, a legislature could be geographically un-representative even if all districts are represented by locals. Conversely, a legislature could be nationally representative if each district is represented by an MP from another district.

a local candidate depend on (1) how fielding local candidates affects a party's vote share (the localness → votes link), and (2) how a change in a party's vote share affects that party's seat share (the votes → seats link) (Cox, 2015). We also consider how electoral institutions shape *voter leverage*, by which we mean voters' ability to translate their preference for locals into electoral outcomes once candidate selection is realised.

3.3.1 Single-member and Multi-member Districts

We expect that the localist vote will be larger – and therefore candidates' localness will matter more for vote shares – in single-member (SM) district elections than in multi-member (MTM) constituencies because the connection between MPs and their constituencies is stronger. SM districts are often thought to foster normative expectations of a close relationship between legislators and the districts they represent: the smaller the constituency a legislator is directly accountable to, the more visible and accessible she will be to voters (Wessels, 1999). The particularistic nature of the electoral linkage in SM districts can therefore be expected to favor a higher degree of candidate recognition and more personalised electoral competition relative to larger, MTM constituencies (Curtice and Shively, 2009; Cain, Ferejohn and Fiorina, 2013). It follows that replacing a non-local candidate with a local candidate in the SM case would affect voting results more than replacing a non-locally representative list with a locally-representative one in the MTM case.

The impact of the electoral system on the translation of votes to seats is more ambiguous. Here we follow Cox, Fiva and Smith (2016), who note an important difference between SM and MTM elections. In a SM district election that is expected to be close, a party might hope to pick up a seat by fielding a local candidate instead of a non-local candidate; in a lopsided SM contest, fielding a local candidate will not affect which party wins the seats, even if it wins some votes. In MTM elections, by contrast, a shift in support has roughly the same probability of changing the allocation of seats regardless of how even or lopsided the expected result is. The impact of a given increment of additional support from 'localists' on the allocation of seats may thus be largest in a competitive SM district, smallest in a 'safe' SM district, and intermediate in a MTM district.[3]

Combining the localness → votes step and the votes → seats step, we conclude that party incentives to field local candidates are highest in competitive SM districts: voters in SM elections respond to localness, and parties competing in tight races are eager to win any marginal support they can. Local representatives may however be very rare in safe seats, where the largest party prefers non-locals and has the ability to 'parachute'

[3]In this sense, we can *define* seat safety in terms of the relationship between the size of the localist vote and the expected margin of victory of the largest party: if the localist vote is large enough to potentially make a difference, the seat is competitive; if it is smaller than the expected margin of victory, then seat is safe.

its preferred candidate: voters in such places may reward a party for running a local, but their marginal support makes no difference to the outcome. MTM districts are an intermediate case: while localists are comparatively fewer here, their votes can never be fully discounted, as even marginal increases in vote share translate into marginal increases in expected seat shares. Therefore, whether an SM system *in aggregate* favors local candidate selection relative to a MTM system would depend in part on the mix of competitive and safe seats in the SM system.

If presented with both local and non-local candidates, voters in an SM election can influence the outcome only if the election is competitive; thus voter leverage is high in the same districts where parties have a strong incentive to field a local candidate. In MTM systems, the extent of voters' leverage is largely dependent on ballot structure. A MTM system that allows preferential voting – for instance, through open lists or multi-winner ranked choice voting – offers a way for voters to express their preference for locals (however slight) within a party list (Renwick and Pilet, 2016; Passarelli, 2020). Preferential voting may thus lead to more geographically representative delegations of MPs than under closed-list systems, where voters cannot change the candidate ranking chosen by the party.[4]

3.3.2 Mixed-member Systems

Mixed-member (MXM) systems offer voters a chance to elect MPs in both a single-member nominal tier and a multi-member list tier (Shugart and Wattenberg, 2001). If we consider these tiers as separate entities, we might expect legislatures in mixed systems to have an intermediate level of geographic representativeness between SM and MTM systems. When we consider possible "contamination effects" between the two tiers (Herron and Nishikawa, 2001; Cox and Schoppa, 2002; Ferrara, Herron and Nishikawa, 2005; Ferrara and Herron, 2005), however, we see reason to doubt that MXM systems are simply a convex combination of SM and MTM systems.

The key contamination effect from our perspective is the tendency for voters to vote for the same party in the multi-member list tier as they do in the single-member nominal tier. Because of this tendency, parties must consider carefully their candidate selection in the single-member tier, as this choice effectively

[4]This stylised framework does not fully account for the range of variation and idiosyncrasies of electoral institutions. Notably, our framework is not ideal to accommodate the US case, where – in spite of generally low-competitiveness SM district elections – parties' control over candidate selection is severely limited by institutionalised primaries. Empirically, as in our sample there is no other case where all parties use primaries for candidate selection, we treat the US as an 'ordinary' type of SM system in the main model (table 3.3). However, it might be possible to think about US-style primaries in terms of a form of preferential voting device in SM systems, where within-party candidate selection is realised *prior to the election* and the choice is limited to only a subset of voters. When we drop the US from the sample in a robustness check of the main model (table C.6 in the Appendix section C), the substantive interpretation of the results is unchanged.

serves as the most 'visible' face of the party in a district, even if they only care about winning seats in the multi-member tier.[5] This contamination effect produces an additional incentive for a party to field a local candidate in the SM tier *independent of seat marginality*: doing so might in fact raise the party's support in the single-member tier *and* the multi-member tier. Furthermore, this contamination effect increases the incentive for a party to field *any* candidate in the single-member tier, even if they do not have much of a chance of winning. The resulting proliferation of candidates in the single-member tier in turn can make these contests more competitive which, as we noted above, further increases party incentives to field a local candidate and the likelihood that the localist vote is decisive in resolving elections in favor of the local candidate.

Contamination effects in MXM systems thus make it more likely that parties field candidates in the nominal tier, that those candidates are locals, and that competition is close enough that the localist vote can be decisive; all of the factors could make the nominal tier of MXM systems more geographically representative than SM systems. The multi-member list tier, conversely, should not present major deviations from what we expect from a MTM system. For instance, the extent to which voters can express their local preference *in the lists* should vary depending on the presence of preferential voting mechanisms in the same way as described for MTM systems.

3.3.3 Summing Up

In this section, we have assessed a number of theoretical considerations to take into account in predicting the probability that a district has a local as its representative (or one of its representatives). Table 3.1 summarises how different types of districts fare according to our criteria of *party incentives* and *voter leverage*. There are a number of countervailing considerations to take into account for each of the types of constituency structure considered, so that in aggregate we have few unambiguous theoretical expectations on their relative performance. The question of electoral systems and spatial representativeness remains therefore primarily an empirical one, which we will seek to address empirically in the following sections. Our theoretical review

[5]There is fairly robust evidence for contamination effects from the nominal tier to the list tier. For instance, Hainmueller and Kern (2008) attribute the overperformance of parties on the list tier in locales where they hold the SMD to the 'spillover effect' of the candidate's incumbency advance. Gschwend, Johnston and Pattie (2003) find that, when German parties field a candidate that over-performs in their constituency vote, they draw more support in the list vote from split-ticket voters in following elections. Ferrara (2004) finds that the placement of SMD candidates affiliated with a party running as part of a pre-electoral coalition boost that *party*'s performance in the PR tier. Qualitative interviews also suggest that parties take cross-tier spillovers into account in crafting their campaign strategies; for instance, according to a Lithuanian party leader quoted in (Ragauskas and Thames, 2020, p. 8): "We allow, actually encourage, [candidates] to personalize the campaign. To campaign for themselves, not even directly for the party, because personal success spills over to list success [...] *if people vote for the SMD candidate, they will also vote for the list.*" [italics added]

		Party incentives to select local candidates		
		Low	Moderate	High
Voter leverage (ability to express a preference for locals)	High		MTM seats with PV (in both MTM and MXM systems)	competitive seats in SM systems SM tier in MXM systems
	Low	safe seats in SM systems	MTM seats without PV (in both MTM and MXM systems)	

Table 3.1: The probability of a district electing local candidates is increasing in (1) party incentives to elect local candidates and (2) voter leverage. PV = preferential vote, SM = single-member, MTM = multi-member, MXM = mixed-member.

does, however, yield some tentative priors on the aggregate level of spatial representativeness of legislatures under different electoral rules:

1. The probability of districts expressing local candidates is most uneven in SM district systems, which combine the most favourable context for local representation in competitive districts with the least favourable setting in safe districts. Therefore, the performance of SM systems should be highly contingent on a 'political' factor relatively independent from the formal features of the electoral systems: levels of seat competitiveness. It follows that, contrary to the 'constituency linkage' argument suggesting better district representation in SM district systems, on this specific measure of spatial representation we do not have reasons to expect them to outperform MTM or MXM systems.

2. Conversely, because of contamination effects, we expect that the single-member district tier of MXM systems *should* make a positive contribution to overall representativeness relative to MTM districts. Not only should there be relatively few 'safe' seats in the nominal tier, but party incentives to select local candidates remain substantial in all single-member districts, regardless of seat marginality. Therefore, MXM systems should yield more representative legislatures than MTM systems, as their additional single-member tier comprises districts where the likelihood of local representation is high. We have no clear expectations on MXM systems' aggregate performance relative to SM systems, as we cannot *a priori* determine the relative impact of the potential sources of spatial bias we find in the two systems:

the multi-member tier in MXM systems, and safe seats in SM systems.

3. Within multi-member districts (both in MTM and MXM systems), the presence of preferential voting mechanisms should increase spatial representativeness of legislatures.

3.4 Cross-country Analysis

We proceed to test these theoretical expectations on a sample of legislatures from 62 democratic countries, for which we measure our Spatial Unrepresentativeness of Legislatures Index (SURLI) from data on MPs' and population spatial distributions. These data are described in subsection 3.4.1, alongside other country-level variables that we use to capture variation in electoral institutions and other socio-political variables in the sample. Subsection 3.4.2 presents the methodology we use to produce our measure of SURLI. Finally, subsection 3.4.3 discusses the results of cross-country analysis assessing how SURLI relates to electoral system features.

3.4.1 Data

Legislator Data

To measure the extent to which legislators' local ties are representative of the population, we gathered legislators' birthplaces at a single point in time for 62 legislatures in democratic countries. We chose to focus on politicians' birthplaces because these are often an important indicator of geographical ties (e.g. the case of the Green Party candidate for Brent Central mentioned above) and because it is by far the most practical measure of legislators' local ties to collect on a broad scale. Ideally, we might assess various other measures of politicians' local ties, such as where they grew up or where they spent most of their adult lives, but this information is not typically available from public sources; one would therefore need to gather this information by directly contacting the legislators or their close associates, which is a challenging data collection project for one legislature, let alone dozens. We thus focus on legislators' birthplaces, which are often reported in their official profiles and Wikipedia pages, and which for many people is an important mark of their geographical identity.

As our source for legislator birthplaces we rely primarily on data from the Global Leadership Project (GLP) (Gerring et al., 2014) and focus exclusively on lower-house members. The GLP dataset we used includes biographical facts about over 38,000 MPs and other top officials in 145 countries, typically as of

2010 or 2011; much of this data was collected from official government websites. Crucially the GLP includes politicians' birthplaces, recorded as text strings. Coverage of birthplaces in the GLP ranges widely, with many cases near 100% coverage and others (including Portugal, Ireland, Jamaica, and Romania) much lower. We undertook substantial effort to collect birthplaces in cases where they were missing in the GLP; we also checked and corrected the variable identifying legislators to distinguish legislators from unelected cabinet officials and party leaders who did not have seats in the legislature. We supplemented the GLP data with original data collection for six other legislatures: Cape Verde, Chile, Cyprus, Macedonia, Montenegro, and Taiwan. After filling in missing entries where possible and editing entries that were insufficiently precise (e.g. referring to regions rather than municipalities of birth), we excluded foreign-born legislators and geocoded birthplaces using the Google Maps API, resolving ambiguities where they appeared. Figure 3.1 lists the countries we analyze and provides information on data completeness for each one: for most legislature-years, we have valid birthplaces for over 90% of legislators.[6] In terms of case selection, we started with the list of all democracies – countries rated free or partly free in Freedom House (2012) – and proceeded to exclude countries for which, between the GLP and our own efforts, we were not able to obtain a sufficient proportion of MP birthplaces.

Population Data

We next need data characterising the spatial distribution of the population, which we will compare to the spatial distribution of legislator birthplaces. Here we face a normative judgment. Should a legislature be considered representative if its legislators' birthplaces match the distribution of citizens' *residences*, or the distribution of citizens' *birthplaces*? The former approach assesses whether people living in different places have the same chance of being represented by a local MP; the latter approach assesses whether people born in different places have the same chance (unconditionally) of becoming an MP. Both approaches have merit, so we gather data to allow both types of analysis. Specifically, we derive two alternative spatial un-representativeness of legislatures measures: one that compares legislators' birthplace distribution to population distribution just before their election (we use 2005 as benchmark year), and one that compares legislators' birthplace distribution to the population distribution in the mean legislator birth year.[7] The dis-

[6] The exclusion of foreign-born legislators is particularly significant for countries like Israel (68% valid entries) and Australia (88% valid entries). In four cases (Argentina, Colombia, Luxembourg, Timor-Leste) we have 2-5 more legislators than the assembly size, so that the share of valid entry is slightly above 100%. This is because we were unable to identify and exclude 'substitutes', who took their seats over the year in which the GLP data was collected.

[7] Legislators' birth year were drawn from the same source as their birthplaces (see previous subsection). The mean legislative birth year ranges from 1951 (France) to 1967 (Macedonia).

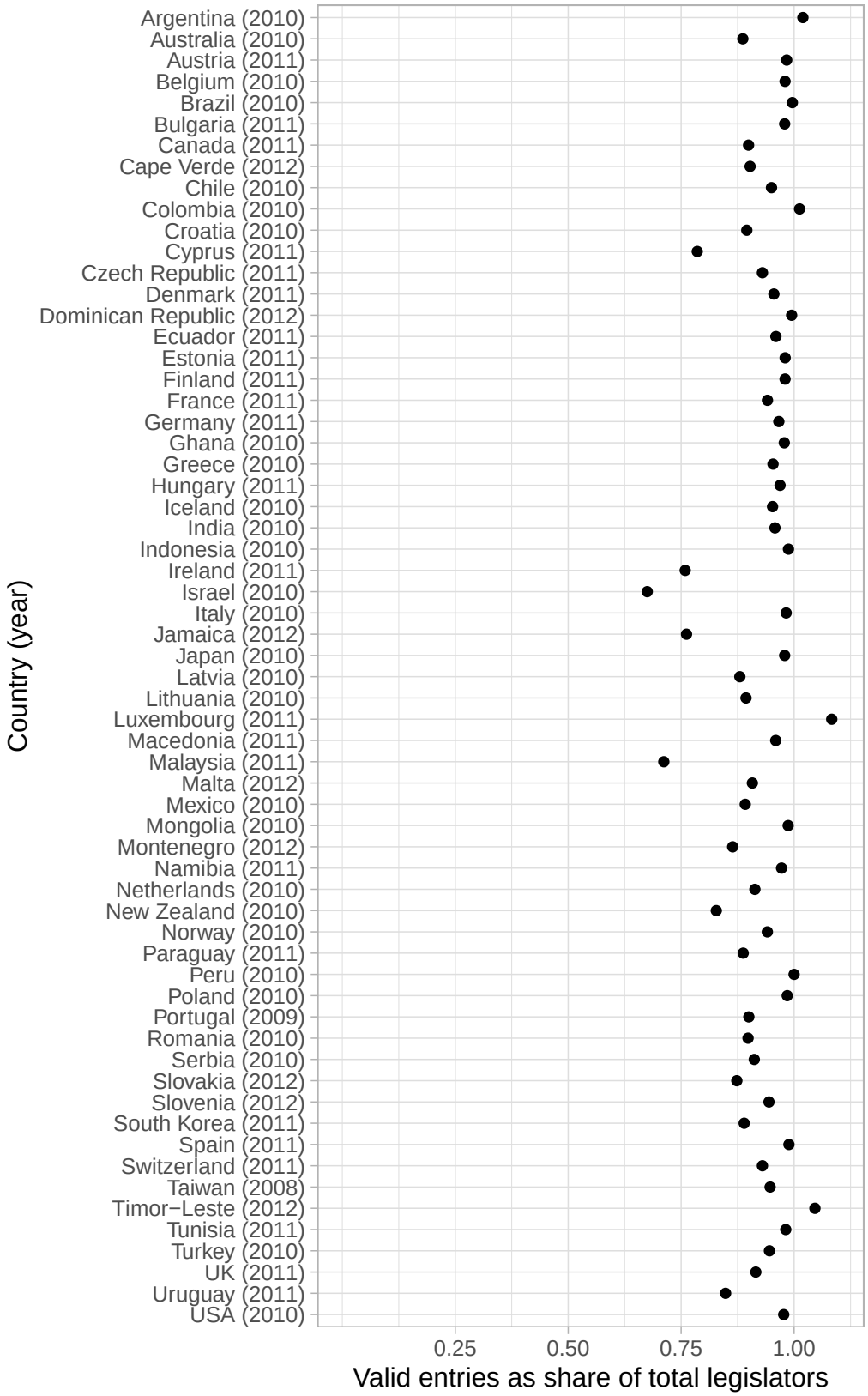


Figure 3.1: Valid non-foreign-born legislator birthplaces as share of assembly size

tribution of population in 2005 may be different from the distribution of population in the mean legislative birth year (and thus the two measures could yield different results) if there are substantial population shifts over time due to e.g. urbanization.

Our data source for population distributions is the gridded population data from the latest version of the History Dataset of the Global Environment (HYDE 3.2), [8] which includes population estimates for every year between 1861 and 2005 at the level of 5-by-5 arcminute spatial grids (i.e. cells of roughly 10km side at the equator). These have subsequently been assigned to countries via geocoding of their centroids, and aggregated up to 15-arcminute side grids. In the aggregation, we made sure that, for each country, these larger cells add up all the 5-arcminute side grids *whose centroid falls in the country*, so as to minimise information loss or measurement error in cells along borders and coastlines.

Country Data

Finally, we gathered country-level variables capturing cross-country variation in electoral institutions in force in the last election prior to legislator data collection, as well as other socio-political characteristics. For *constituency structure*, we constructed both a categorical variable (single-, multi- and mixed-member systems) and an interval variable measuring the share of legislators elected in multi-member seats, which we use in the cross-country regression alongside a 'mixed-member system' dummy. A dummy variable for *preferential voting* was coded via a qualitative assessment of ballot rules from legislation or secondary sources. The variable takes the value of 1 if (1) voters can express a preference for one or more individual candidates that is functionally different from a party vote, and (2) the preference vote can practically determine at least in part the allocation of seats,[9] and 0 otherwise. We also introduce in our models a measure of *district magnitude*, to account for the fact that for very large districts a possible source of spatial inequality in representation might be the unequal selection of candidates *between different parts of the same district*. To record this variable, we followed the coding rules in Carey and Hix (2011), and computed *median district magnitude* and *mean district magnitude* variables, calculated excluding from the count constituencies that are either non- or extra-territorial (e.g. ethnic constituencies, Greenland, nationals abroad). We also collected country-level data on *level of democracy* (V-Dem's unified democracy score), *federalism* (using the list of countries in Roeder, 2009) and a measure of geographic economic inequality, the *spatial Gini* (population-

[8] https://www.pbl.nl/en/publications/new-anthropogenic-land-use-estimates-for-the-holocene-hyde-32

[9] This second criterion excludes, for instance, the case of Norway, where voters can in theory rearrange the ranking of candidates or cross out candidates they do not want elected, but more than 50% of the party voters have to 'move up' or 'cross out' a candidate to alter the party's preferred ranking – which has never happened, making these options entirely symbolic (Aardal, 2007).

	Mean	SD	Min	Max	N
Share multi-member districts	0.758	0.389	0	1	62
Median district magnitude	20.49	44.019	1	250	62
Mean district magnitude	22.927	45.449	1	250	62
Preferential voting	0.419	0.497	0	1	62
Population ('000)	51,989	165,073	319	1,250,288	62
Population (log)	16.20	1.737	12.67	20.95	62
Km^2 Land area	917,435	2,125,164	320	9,147,420	62
Km^2 Land area (log)	11.952	2.030	5.768	16.029	62
GDP per capita (2011 $ PPP)	24,764	22,435.250	1410	105,265	62
GDP per capita (log)	9,658	1.045	7,252	11,564	62
Level of democracy	1.145	0.555	0.160	2.263	62
Federalism	0.226	0.421	0	1	62
Spatial Gini	0.0492	0.027	0.0128	0.0623	60
Constituency structure					
Single-member	-	-	-	-	9 (16.1%)
Multi-member	-	-	-	-	43 (69.3%)
Mixed-member	-	-	-	-	10 (14.5%)

Table 3.2: Country-level data

weighted Gini index of estimated regional GDP per capita, estimated from satellite nighttime light data in Lessmann and Seidel, 2017). Finally, we also collected data on *population*, *land area* and *GDP per capita* from the World Bank's (2010) World Development Indicator catalogue. Descriptive statistics for these variables are presented in Table 3.2.

3.4.2 Measurement of SURLI

To measure the Spatial Un-Representativeness of Legislatures Index (SURLI) for a country's legislature, we begin with the geocoded location of each MP's birthplace and the proportion of the population in each grid square of the country. After assigning legislators' birthplaces to grid squares, we can express the two distributions (birthplaces and population) in terms of the proportion observed in each grid square. Our objective is to measure the discrepancy between these two distributions in a way that is comparable across countries of greatly differing territorial size and shape. A natural choice for comparing the two distributions is the Earth Mover's Distance (EMD, Rubner, Tomasi and Guibas, 2000), a metric borrowed from computer science and introduced to political science by Lupu, Selios and Warner (2017). Simply put, EMD measures the amount of work (mass times distance) necessary to transform one distribution to another; as Lupu, Selios and Warner (2017) argue, this closely matches our intuitions about when one distribution is close to another. In our case, EMD measures the minimum total amount of travel necessary to move an equal

number of citizens to each MP birthplace.

Although EMD should pick up variation in the (un-)representativeness of a legislature due to e.g. the electoral system, in cross-country comparisons it will also reflect other differences between countries that may obscure these patterns. Notably, it will depend on a country's size: if a country consists of just two cities separated by a desert, then (assuming the proportion of MPs born in each city differs to some extent from the proportion of people living there) the EMD is increasing in the distance between the cities; also, if a country consists of a single grid square, its EMD is zero regardless of its political institutions. The EMD will also depend on the size of the legislature relative to the size of the country: generally, the more seats in the legislature the more closely the distribution of MP birthplaces can potentially match the distribution of inhabitants. Clearly, an investigation of the effect of electoral institutions on geographical representation using observational data requires addressing these country-specific determinants of the EMD score. As a first line of defence we control for land area and population in the regressions below. To further address these and other sources of heterogeneity, we compute the SURLI for each country as a z-score that compares the country's actual EMD against the distribution of EMDs we obtain for that country when MPs are selected at random from the population over a large number of simulated draws. Thus a positive SURLI for a country means that the country's observed EMD is higher than the average EMD across random representative legislatures for that country; country i's SURLI will be higher than country j's if country i's EMD is higher *relative to its own null distribution* than country j's is.[10]

Unfortunately, the time to compute EMD increases exponentially in the number of grid squares, so that computing SURLI for a large country like the US (with over 14 thousand grid squares) can take weeks. The complexity of computing EMD in two or more dimensions is well known, prompting efforts to develop efficient implementations and approximations (e.g. Cuturi, 2013). We discovered that, at least for our application of the algorithm on square grids, we could obtain a very efficient approximation to the EMD by computing the one-dimensional EMD for each of several rotations of the gridded map (e.g. east to west, northeast to southwest, north to south, northwest to southeast) and averaging those. As we show in section A of the appendix, the resulting estimate correlates very highly with the two-dimensional EMD in actual cases and can be computed dramatically faster. In one dimension, in fact, the EMD between two distributions is known to be equivalent to the computationally cheap procedure of integrating the area between two cumulative distribution functions (CDFs) (as proven in Cohen and Guibas, 1997, pp. 13–16), which is the method

[10]More formally, let d denote a country's actual EMD, and let $\delta = \{\delta_1, \delta_2, \ldots, \delta_M\}$ denote M counterfactual EMDs assuming a representative legislature. Then SURLI is $(d - \overline{\delta})/\mathrm{sd}(\delta)$.

Golder and Stramski (2010) had suggested for comparing distributions.[11]

Thus SURLI is computed in three steps. First, for each country we measure the spatial difference between the distribution of the population and the distribution of MP birthplaces (both allocated to squares on a gridded map) by computing the one-dimensional EMD (i.e. CDF discrepancy) in each of four rotations of the map and averaging these. Second, for each country we recompute the same measure but for 500 fictional legislatures where MPs are drawn at random from the population distribution. Third, we compute SURLI as the z-score of the value obtained in the first step over the distribution obtained in the second step, which provides a value that should be comparable across countries of different sizes. Note that SURLI scores can take negative values, indicating that the actual measure of discrepancy between MPs' birthplaces and the population distribution is *smaller* than we would expect if legislators were chosen at random. The output of this measurement exercise for all countries in our sample is plotted in section B of the appendix.

[11] Lupu, Selios and Warner (2017), in advocating EMD over Golder and Stramski (2010)'s method, were apparently unaware of the equivalence. Our contribution in section A of the appendix is to show that one can closely approximate a two-dimensional EMD by repeatedly integrating one-dimensional CDF discrepancies.

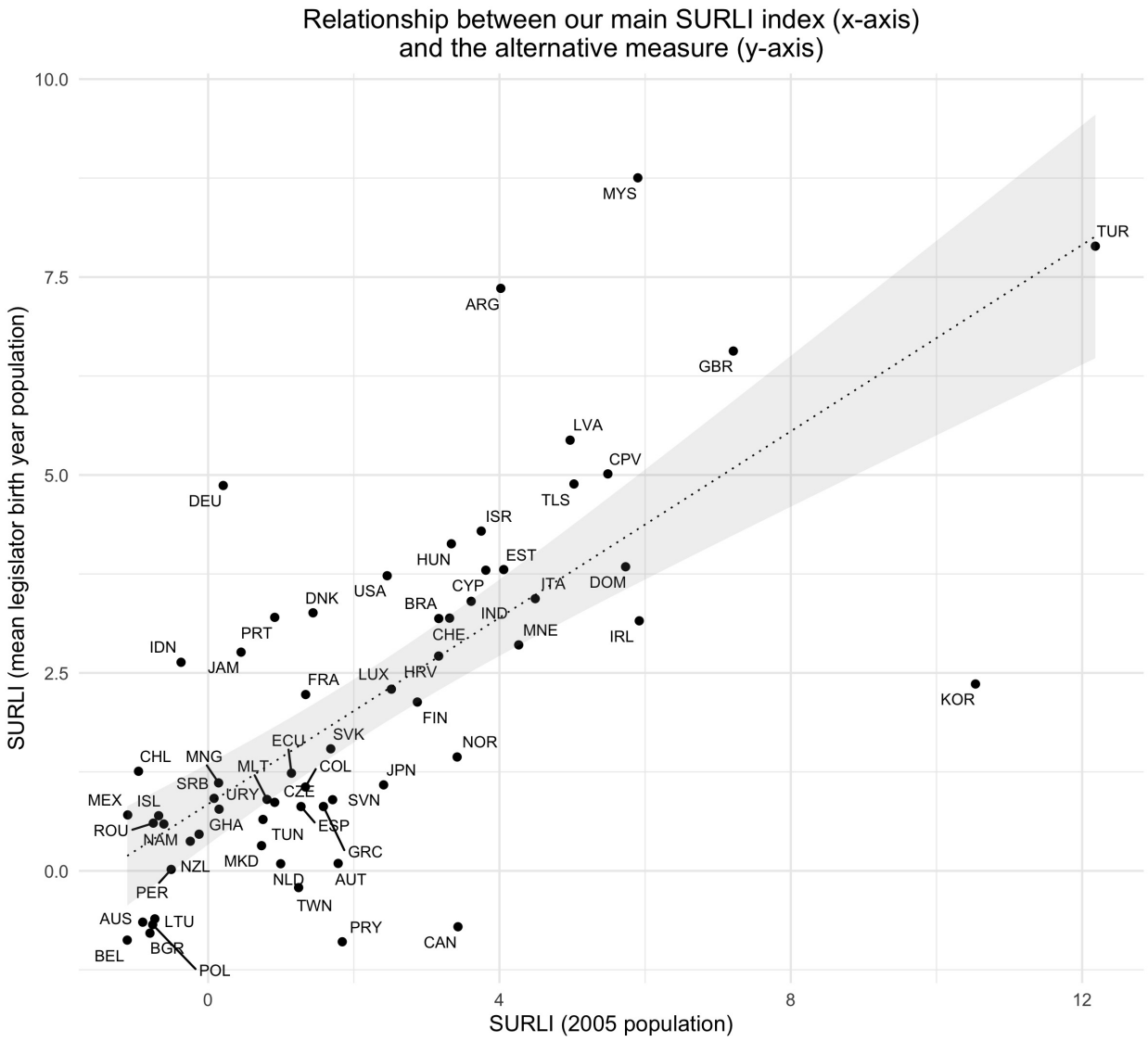


Figure 3.2: Comparison between SURLI computed against the distribution of the population around the time of the election, and SURLI computed against the distribution of population in the mean birth year of a country's legislators.

Additionally, we repeat the same procedure with an alternative target population distribution – population distribution in the mean legislator birth year – which serves as a proxy for the distribution of birthplaces of the potential candidate population. We choose to derive this *SURLI (mean legislator birth year)* alternative set of scores to check our results against an index that aims to net out the effects of e.g. internal migration. The intuition is simple: if the population in one part of the country has grown faster than the others over the average lifespan of legislators, it will have relatively more voters than it has birthplaces of

potential MPs; this will bias upwards the overall SURLI score. The two resulting measures of geographic un-representativeness of legislatures capture two slightly different dimensions of spatial bias. By comparing legislators' birthplaces to population distribution around the time these legislators were elected, the main variable *SURLI (2005 population)* measures the extent to which each part of the country is represented by a 'local-born' legislator, regardless of the different share of local-born people in each territorial unit. By comparing legislators' birthplaces to population distribution around the time these legislators were born, the variable *SURLI (mean legislator birth year distribution)* measures the extent of inequality of access to public office between people born in different parts of the country. Both these dimensions – we believe – are substantively interesting. As shown in figure 3.2, the correlation between the two sets of scores is quite high (Pearson's $r = 0.72$), though there are substantial differences for some countries (particularly those with significant internal migration due to urbanization or, in the case of Germany, reunification). The correlation between land area and the main measure of SURLI is -0.03, suggesting that our method does appear to net out differences in country size, as desired.

3.4.3 Results and Discussion

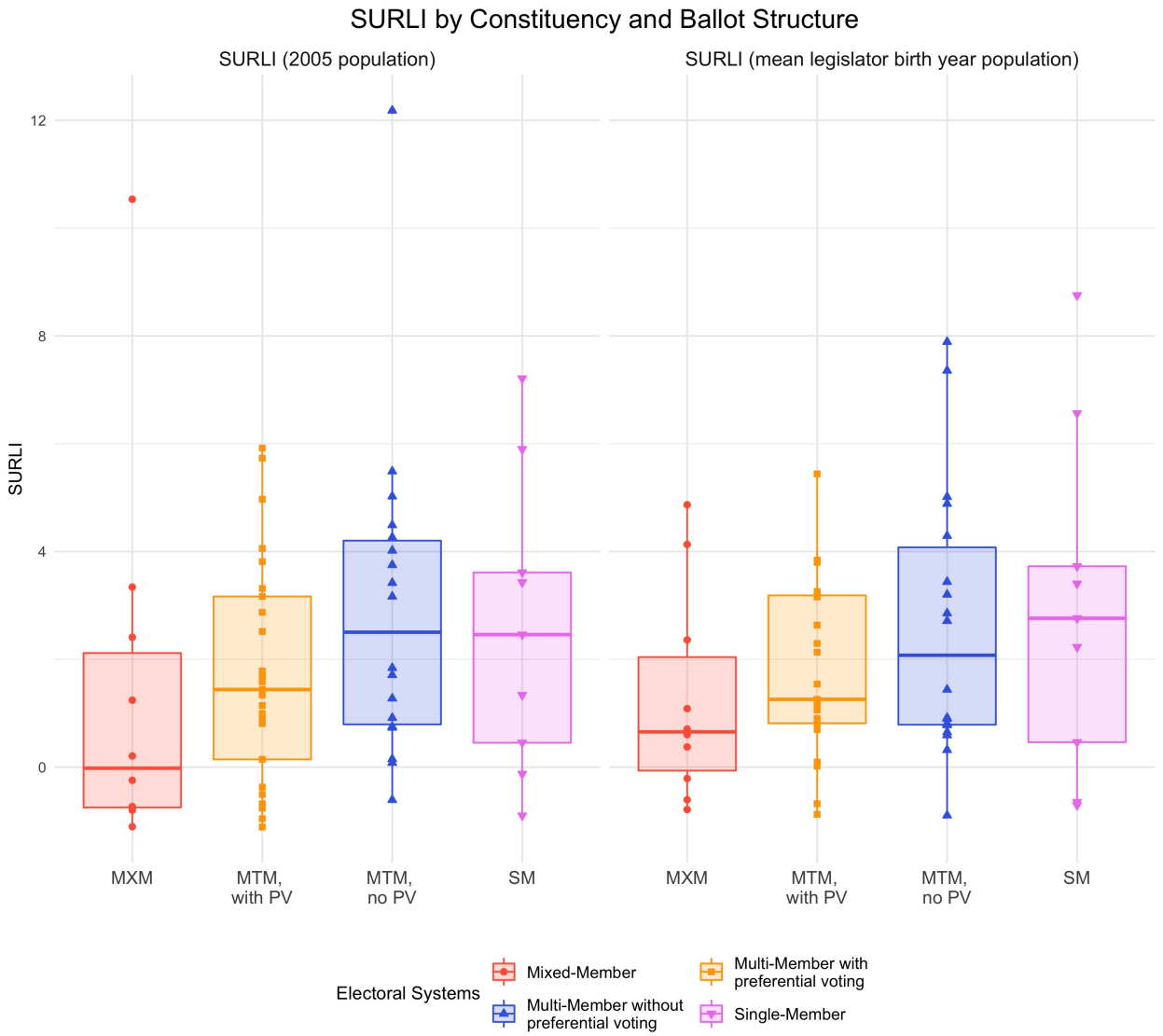


Figure 3.3: SURLI scores by electoral systems. SM = single-member, MTM = multi-member, MXM = mixed-member, PV = preferential voting. The only MXM country in our sample with PV in the MTM tier (Lithuania) was grouped with MXM for illustrative purposes.

Having derived the spatial un-representativeness of legislatures index (SURLI), we can now proceed to investigate its relationship with electoral institutions. Figure 3.3 shows the distribution of *SURLI (2005 population)* and *SURLI (mean legislator birth year population)* across four major families of electoral systems, defined by their constituency and ballot structures. For both versions of SURLI, the median value in MXM systems is the lowest, with scores below 1 indicating that the spatial bias in the median MXM system is less

than one standard deviation higher than the spatial bias we would expect from random draws of legislators from the distribution of the target population. Preferential voting in MTM districts is associated with slightly lower median SURLI scores, and SM districts fare poorly relative to MTM districts. Conversely, the scatterplots in Figure 3.4 show no discernible relationship between SURLI and district magnitude for either SURLI measure.

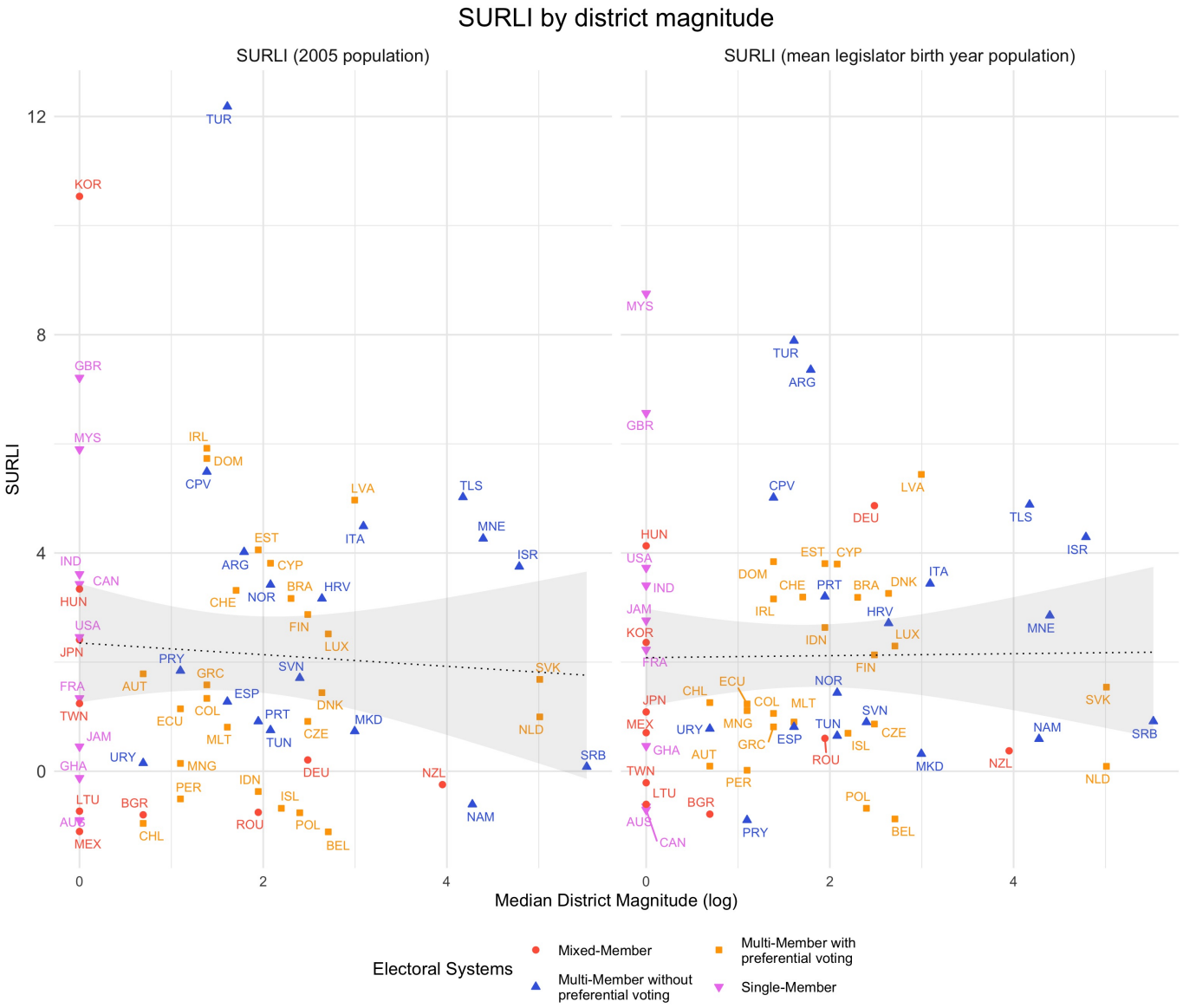


Figure 3.4: SURLI scores by median district magnitude.

Table 3.3 show the results of OLS regression models where SURLI is regressed on a series of electoral system, demographic and institutional characteristics of our country-sample. In models 1 and 3, we op-

erationalise constituency structure as a three-factor categorical variable; in models 2 and 4, we employ a continuous variable for the share of legislators elected via MTM districts alongside a dummy for mixed-member systems. This dummy effectively captures the 'additional' effect of mixed-member relative to the predicted value of SURLI if these systems simply functioned as hybrids of single- and MTM systems. *SURLI (2005 population)* is used as dependent variable in models 1 and 2; the alternative specification computed against the proxy for birthplace distribution of the potential population of office-seekers – i.e. *SURLI (mean legislator birth year population)* – is used in models 3 and 4. We control for *preferential voting*, the log of *median district magnitude*, the log of *population*, the log of *land area* and the log of *GDP per capita* and *democracy score*.

	Dependent variable:			
	SURLI (2005)		SURLI (mean MP birth year)	
	Model 1	Model 2	Model 3	Model 4
Constituency Structure 1[a]				
Multi-Member	2.63** (1.15)		1.63* (0.96)	
Single-Member	1.56 (1.28)		2.19** (1.07)	
Constituency Structure 2				
Mixed-Member		−2.08** (1.01)		−1.92** (0.84)
Share Multi-Member		0.64 (1.34)		−0.35 (1.12)
Preferential Voting	−1.58* (0.82)	−1.46* (0.81)	−0.82 (0.68)	−0.88 (0.68)
Median dist. mag. (log)	−0.40 (0.30)	−0.35 (0.29)	−0.01 (0.25)	−0.03 (0.25)
Population (log)	0.39 (0.32)	0.37 (0.32)	0.29 (0.27)	0.30 (0.27)
Land area (log)	−0.43* (0.25)	−0.43 (0.26)	−0.28 (0.21)	−0.28 (0.21)
GDP p.c. (log)	1.10* (0.56)	1.13* (0.56)	0.68 (0.47)	0.66 (0.47)
Democracy score	−1.73 (1.09)	−1.82* (1.09)	−1.55* (0.91)	−1.51 (0.91)
Constant	−8.31 (5.49)	−6.46 (5.50)	−5.12 (4.60)	−3.07 (4.59)
Observations	62	62	62	62
R^2	0.21	0.20	0.18	0.18
Adjusted R^2	0.09	0.08	0.05	0.05
Residual Std. Error (df = 53)	2.58	2.59	2.16	2.17
F Statistic (df = 8; 53)	1.74	1.68	1.44	1.42

[a]: reference category: Mixed-member; *p<0.1; **p<0.05; ***p<0.01

Table 3.3

The results in Table 3.3 indicate that, even after including control variables, legislatures tend to be less geographically unrepresentative in mixed-member systems than elsewhere. Models 1 and 3 indicate that SURLI in MXM systems is on average between 1.6 and 2.6 points lower than in MTM systems and between 1.6 and 2.2 points lower than SM systems (depending on the SURLI measure), controlling for district magnitude, population, land area, GDP per capita, and the democracy score; three of the four estimated differences are significant at the .1 level or lower. Models 2 and 4 indicate that SURLI in mixed-member systems is between 1.9 and 2.1 points lower than the rest of the sample, net of the share of legislators elected via MTM districts; both estimates are significant at the .05 level. The findings for *preferential voting* are less clear-cut: the

difference is in the predicted direction in all models – i.e. lower SURLI for systems with preferential voting – but it only reaches significance at $p < 0.1$ in models 1 and 2, where SURLI (2005) is the outcome. We find no significant conditional association between *median district magnitude* and either dependent variable. Larger countries seem to have, *ceteris paribus*, more representative legislatures than smaller ones; this could be due to systematic differences in spatial bias across these settings, but it could also be that SURLI as we compute it simply cannot be very high in countries with very few grid squares.[12]The democracy score tends to be negatively associated with SURLI, which suggests that descriptive spatial representativeness at least partly maps onto other dimensions of democratic quality. In section C of the appendix, we present alternative specifications where *federalism* and *spatial Gini* are substituted to *democracy score*, and *mean district magnitude* is employed in lieu of the median: the substantive interpretation of the main coefficients of interest is unchanged. The negative relationship between mixed-member constituency structure and SURLI is also robust to the introduction of a presidentialism dummy, a magnitude $\times$ preferential voting dummy, a dummy recording whether there are legal residency requirements for candidates, as well as to alternative specifications of the main models where assembly size is used instead of population and number of districts instead of district magnitude.

In sum, in the cross-country analysis we find no evidence that single-member district systems have more geographically representative legislatures than multi-member district systems. Whatever the advantages of SM constituencies for maintaining connections between MPs and localities, SM systems do not seem to engender more equal access to political office for people born in different parts of the country. Instead, MXM systems appear the most conducive institutional environment for geographically representative legislatures. We do not find a notable association between district magnitude and SURLI, but there is some tentative evidence that mechanisms allowing voters to express a candidate preference alongside a party preference may improve spatial representativeness of legislatures.

3.5 Within-country Analysis

The cross-country analysis suggests that mixed-member systems are not only superior to multi-member districts on our measure of spatial representativeness, which is in line with our theoretical expectations; *but they also outperform SM districts.* In this section, we use within-country analysis to further investigate

[12]To make sure this does not affect the interpretation of our coefficients of interest, we present in Appendix C, table C.5 the main model excluding micro-countries from the sample.

what makes mixed-member systems more geographically representative. In Section 3.3 we argued that contamination effects in mixed systems could lead to better local representation in the nominal tier of mixed systems than in single-member systems; essentially, this is because parties in mixed systems expect to gain in the multi-member tier by fielding attractive candidates in the nominal tier, which means that more candidates are local (even in safe seats) and fewer seats are safe. In our within-country analysis we can test this argument. Rather than using SURLI (which applies to a whole legislature), we look to the district level to see whether each district elects a local MP and how this relates to competitiveness differently in single-member and mixed systems.

For reasons of data availability, we select two countries – Britain and Germany – as examples of, respectively, a single-member district plurality and a compensatory mixed-member system. The analysis is restricted to Germany's single-member tier. In sum,

1. We expect German single-member district elections to be more competitive than British single-member district elections, because contamination effects induce entry.

2. Compared to British MPs, we expect a greater proportion of German MPs elected from the nominal tier to be local to the seat they represent, for the reasons stated above.

3. In the UK, we expect the probability of a local MP to be lower in safer seats. We expect this to be less the case in the German nominal tier due to contamination effects: a party that expects to win easily may still field a local candidate to boost its results in the multi-member tier.

Given the many ways in which the UK and Germany differ politically, one should not expect a comparison of the two cases to produce conclusive evidence for or against any theory. Patterns consistent with the above expectations would, however, tend to corroborate our interpretation of the broader patterns we find.

We created for each country a dataset that combines biographical information on legislators, district-level data on parties' electoral performance, and spatial data from constituency boundary digital vector files. For both countries, our primary source for MPs' data is the `legislatoR` database, which includes – among other information – legislator birthplaces, party affiliation, constituency and electoral tier. We selected British and German single-member district MPs elected in the past six parliamentary elections (respectively, 2001-2019 and 1998-2017), for a total of 3,971 British and 1,823 German legislator-session entries.[13] We complemented the birthplace data in the dataset with further research and geocoded the

[13]The British sample also includes MPs elected via by-elections held between 1997 and 2019.

locations found, yielding virtually complete coverage of the sample for this variable (see the first row of tables 3.4 and 3.5). Using constituency names and codes, we linked each entry to party shares of the vote *in the previous election* (this approximates parties' priors in the candidate selection stage). To account for redistricting, for the UK sample we combined data from the House of Commons library with notional seat shares estimated by Rallings and Thrasher;[14] for Germany, we used both 'real' and notional district-level results published for each election by the *Bundeswahlleiter* (German federal electoral commission). Constituencies were then linked to geocoded vector polygon data in shapefiles obtained from the *UK Data Service* and the *Bundeswahlleiter*.

	United Kingdom						
Election year	2001	2005	2010	2015	2017	2019	overall
% valid birthplaces	0.93	0.99	0.98	0.98	0.98	0.97	0.97
Mean margin in last election	0.24	0.23	0.19	0.18	0.24	0.24	0.22
Med. margin in last election	0.21	0.20	0.18	0.17	0.24	0.23	0.20
% Safe seats (> 10% margin)	0.76	0.79	0.74	0.69	0.81	0.74	0.76
% Ultrasafe seats (> 20% margin)	0.51	0.51	0.43	0.43	0.59	0.57	0.51
Med. distance MP birthplace-seat (km)	93.97	100.20	89.21	73.09	72.00	57.74	79.50
% MPs born in seat	0.25	0.24	0.25	0.28	0.29	0.32	0.27

Table 3.4: Descriptive statistics, UK MPs sample

	Germany (single-member district tier)						
Election year	1998	2002	2005	2009	2013	2017	overall
% valid birthplaces	0.99	0.99	0.99	1.00	1.00	1.00	1.00
Mean margin in last election	0.14	0.13	0.15	0.14	0.14	0.18	0.14
Med. margin in last election	0.11	0.10	0.13	0.12	0.11	0.16	0.12
% Safe seats (> 10% margin)	0.55	0.53	0.59	0.57	0.54	0.68	0.57
% Ultrasafe seats (> 20% margin)	0.27	0.22	0.27	0.23	0.26	0.39	0.27
Med. distance MP birthplace-seat (km)	29.48	26.16	24.83	20.68	18.54	18.88	21.76
% MPs born in seat	0.62	0.71	0.70	0.73	0.78	0.74	0.71

Table 3.5: Descriptive statistics, German single-member district MPs sample

We address the first two hypotheses – concerning district competitiveness and extent of local representation – via simple descriptive analysis. From real or notional district-level electoral data, we computed a *margin in last election* variable, as the difference between the share of the vote for the largest party in the previous election and the share of its closest competitor. Furthermore, we used the digital vector data to compute, for each legislator, the seat in the current election that includes the legislator's birthplace. Additionally, we computed the geodesic distance between the legislator's birthplace and the centroid of the seat

[14] In the period under consideration, English and Welsh constituency boundaries changed between the 2005 and 2010 elections, while Scottish and Northern Irish seats were redistricted between 2001 and 2005. We could not find notional estimates of party shares in the 2001 election for Northern Ireland seats as configured in 2005. Data for 2005 notional results are available from Pippa Norris's personal website at https://www.pippanorris.com/data; we are thankful to Lewis Baston for providing data for 2001 notional results for Scotland.

	Dependent variable:	
	MP born in seat (UK)	
	Model 1	Model 2
Party margin in previous election	-0.896^{**} (0.444)	-1.096^{**} (0.488)
Constituency area (km^2	-0.0001 (0.0001)	-0.0002^{*} (0.0001)
Party[a]		
Labour	1.205^{***} (0.201)	2.468^{***} (0.820)
Lib Dem	0.354 (0.348)	2.059^{**} (0.958)
Other	1.280^{***} (0.379)	2.235^{**} (1.038)
SNP	0.736^{**} (0.312)	1.963 (1.382)
Election[b]		
2005	-0.034 (0.328)	0.870 (0.826)
2010	0.206 (0.297)	1.133 (0.763)
2015	0.658^{**} (0.314)	1.778^{**} (0.778)
2017	0.554 (0.340)	1.816^{**} (0.834)
2019	0.809^{***} (0.312)	1.960^{**} (0.762)
By-Election	0.192 (0.658)	1.574 (1.368)
Constant	-1.544^{***} (0.285)	-2.536^{***} (0.733)
Party × Election Interaction	No	Yes
Observations	864	864
Log Likelihood	-511.192	-497.661
Akaike Inf. Crit.	1,048.384	1,065.323

[a] = ref. cat. Conservative, [b] = ref. cat. 2001
*p<0.1; **p<0.05; ***p<0.01

Table 3.6: Binomial logistic models. The estimates capture the variables' effect on the probability that a newly elected British MP is born in the constituency she represents. Model 2 includes Party × Election interactions: interaction terms' coefficients not shown for reasons of space.

	Dependent variable:	
	MP born in seat (German SM tier)	
	Model 1	Model 2
Party margin in previous election	1.235^{*} (0.692)	1.661^{**} (0.811)
Constituency area (km^2)	-0.0002^{**} (0.0001)	-0.0002^{*} (0.0001)
Party[a]		
Others	0.395 (0.528)	13.942 (882.744)
SPD	0.088 (0.231)	−0.066 (0.565)
Election[b]		
2002	0.500^{*} (0.299)	0.394 (0.601)
2005	0.456 (0.342)	0.267 (0.617)
2009	0.734^{**} (0.318)	0.737 (0.574)
2013	0.890^{**} (0.365)	0.621 (0.593)
2017	0.282 (0.336)	−0.169 (0.575)
Constant	0.465 (0.289)	0.571 (0.537)
Party × Election Interaction	No	Yes
Observations	604	604
Log Likelihood	−365.319	−361.198
Akaike Inf. Crit.	750.638	760.396

[a] = ref. cat. CDU/CSU, [b] = ref. cat. 1998
*p<0.1; **p<0.05; ***p<0.01

Table 3.7: Binomial logistic models. The estimates capture the variables' effect on the probability that a newly elected German single-member district MP is born in the constituency she represents. Model 2 includes Party × Election interactions: interaction terms' coefficients not shown for reasons of space.

she represents. Finally, we combined these two pieces of information to create a binary *MP born in seat* variable that takes the value of 1 if either (1) the legislator's birthplace falls within the seat she represents, or (2) the legislator is born within 20km of the centroid of the seat she represents. This double-safe coding rule is meant to minimise type II measurement errors in districts of varying size. For instance, using only the first criterion, any legislator born in London would be coded as being from the central seat of 'Cities of London and Westminster', but obviously given that information they are just as likely to be 'local' to any other London seat. Equally, using only the second criterion an MP born in Dumfries – the major settlement in the 4,000 km^2 rural Scottish constituency of Dumfries and Galloway – would not be coded as local because the town is not close enough to the constituency's centroid.

The measures of district competitiveness and local representation we provide in Tables 3.4 and 3.5 are consistent with our priors: UK elections are less competitive than German SM district races, and British MPs are far less likely to be born in their constituencies than German SM district legislators. The most competitive election in the UK in terms of mean and median previous margin (2015) is less competitive than the least competitive election in German SM district seats (2017). Over three quarters of UK constituencies are 'safe' (the margin is larger than 10%) and over half are 'ultra-safe' (the margin is larger than 20%), against 57% and 27% in German single-member districts. As far as local representation is concerned, the contrast is equally stark. Under the aforementioned definition of being 'born in the seat', 71% of German single-member district legislators are local to their area, against only 27% of British MPs. This does not appear to be simply an artifact of German constituencies being larger: the median distance between an MP's birthplace and the centroid of the constituency she represents is 22km in Germany and 79km in the UK.

Next we probe the relationship between seat marginality and parties' incentives to parachute candidates in the two cases. We first restrict the two samples to newly-elected legislators only, in order to have observations that capture outcomes of party and voter choices taken at the same time and based on known priors of seat competitiveness.[15] Within these samples, which amount to 904 entries for the UK and 605 for Germany's single-member district tier, we model the binary outcome variable *MP born in seat* as a function of how well the MP fared in the same seat at the previous election. This *party margin in previous election* regressor is computed as the difference between the vote share of the winning MP's party in the previous election and the vote share of the top-ranking competitor. Unlike the absolute measure of marginality discussed before, this

[15]Incumbents may have secured a seat many electoral cycles prior to each individual election, when the calculus of party selection was different: e.g. a seat may have 'become' more or less safe over time. But parties seldom reassess candidate selection for seats in which a legislator intends to run again, demanding her to step down for a candidate they like better if the seat has become safer, or for a candidate with stronger local credentials if the seat has become more competitive.

variable can take negative values when the legislator's party lost the previous contest (or at least would have, under the current boundaries). As controls, we include party and election dummies as well as a measure of constituency land area (a proxy for urban/rural seat distinction). Moreover, we present an alternative specification (model 2) including interactions between party and election dummies, so that comparisons are effectively between safer and less safe MPs from the same party in the same year. In section D of the Appendix we present alternative specifications of the models, where we employ a probit link function, use the log transformation of the constituency area variable, and introduce controls for country or region and MP's gender.

The logistic regression coefficients for the British and German samples of legislators are shown in tables 3.6 and 3.7; the marginal effects of party margin on the dependent variable across model specifications are plotted in Figure 3.5. In the UK sample, we find the expected inverse relationship between seat safety and likelihood of electing a legislator born in the constituency: MPs elected from safer seats are less likely to be born in the district than those elected from less safe seats, in part because party leaders use these seats as comfortable destinations for insiders. Strikingly, the relationship is instead *positive* in German single-member districts: MPs elected from safer seats are *more* likely to be born in the district than those elected from less safe seats. This may be because the opportunity to win party-list votes through spillovers is greater in party strongholds, or because potential local candidates are more plentiful in such places.

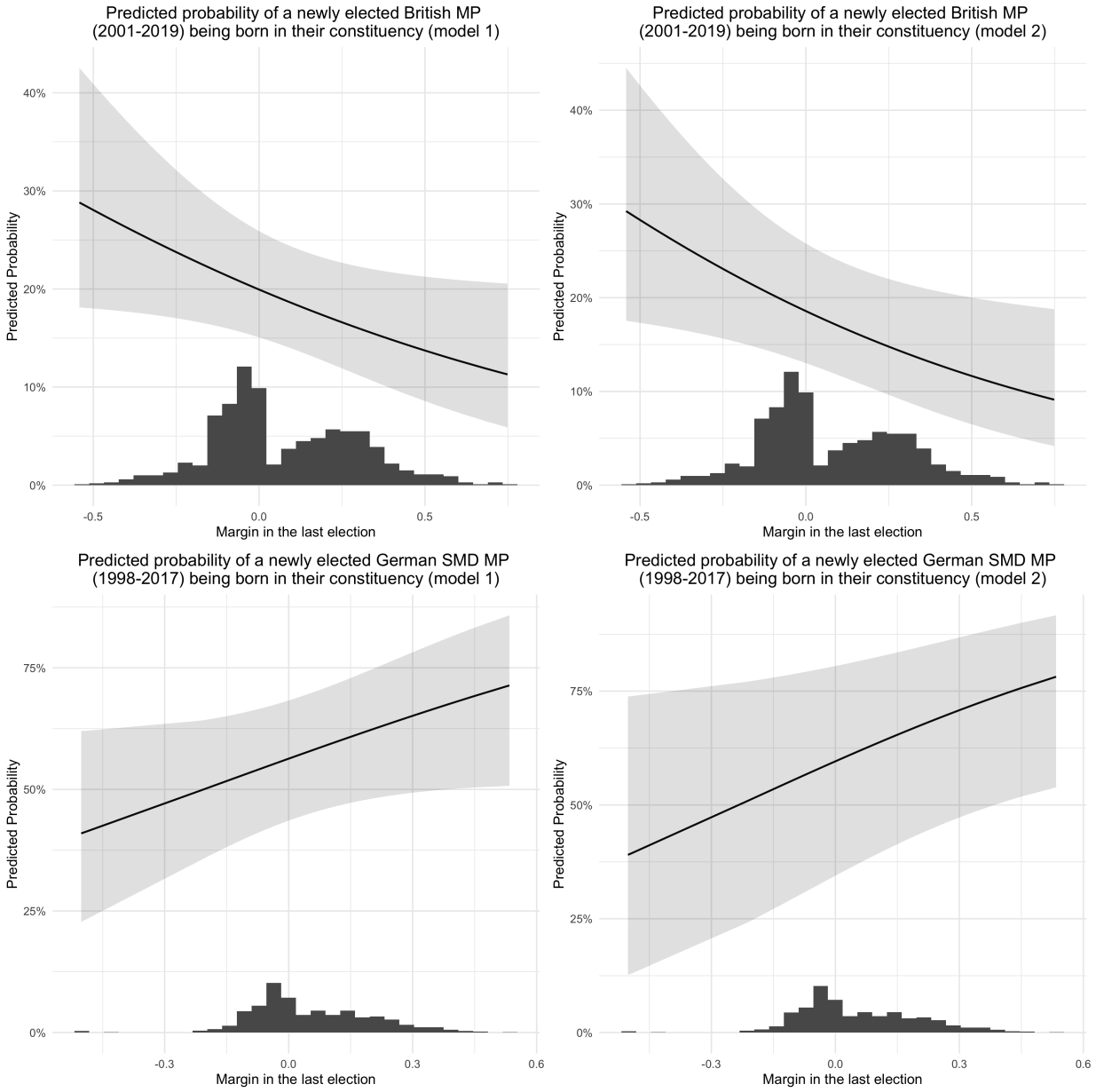


Figure 3.5: Marginal effects of seat safety on the probability that a newly elected single-member district MP is 'local' to her seat (logistic models in Tables 3.6 and 3.7).

3.6 Conclusion

In the title of their influential edited volume, Shugart and Wattenberg (2001) asked whether mixed-member systems offer the 'best of both worlds' relative to single-member districts and proportional representation. From the point of view of descriptive representation of places in parliaments, our analysis suggests that we

can answer that question positively. Judging by our new measure, legislatures elected under mixed-member systems seem to be significantly more geographically representative than those elected in countries with other constituency structures. This finding chimes with our theoretical intuition that, due to contamination effects, the single-member tier of mixed-member systems is more conducive to electing local candidates than either single-member districts in single-member systems or multi-member districts. Alongside a cross-country analysis, we provide additional evidence in this direction by comparing German legislators elected in single-member districts with British MPs. We found that a much larger proportion of these German MPs were born in their district; moreover, safer seats are less likely to be represented by a local in Britain, whereas the opposite is true in Germany.

As a first stab at an understudied dimension of representation, the analysis presented has important limitations. Our cross-country analysis uses a rather coarse system to classify electoral institutions. We mentioned in passing the possible effect of US-style party primaries, but there are other institutional features that we just do not have enough variation in the sample to study comparatively at this stage: different electoral formulae in SM systems, the nature of tier linkage in mixed-member systems, legal electoral thresholds in multi and mixed-member systems etc. An additional limitation, already noted in section 3.4.1, is that inferring MPs' ties to geographic locales from birthplaces discounts other aspects of a legislator's biography – education, work, length of residence – that may link her to a community. Systematically gathering this information for a large set of countries is an enormous task, but the resulting measure might better characterise MPs' geographical ties. Another direction of research might involve devising a measurement for the *substantive* representation of places rather than their *descriptive* representation.

Nonetheless, we believe that this paper makes valuable contributions in at least two senses. From a methodological point of view, it develops a practical method for measuring the congruence between spatial distributions in a comparable way across polities and can therefore be applied to the study of other spatial inequalities in political outcomes. From a substantive point of view, the analysis has normative implications for electoral system design, providing evidence against the 'constituency linkage' argument according to which SM districts lead to better, more personalised representation of locales: an argument not only made by supporters of existing majoritarian systems (Kelly, 2008) but also often conceded by electoral reformers (Jenkins, 1998; British Columbia Citizens' Assembly on Electoral Reform, 2004). To the extent that having legislators that reflect the geographic diversity of a country can be considered a relevant descriptive dimension of local representation, it emerges that other electoral institutions – mixed-member systems, and possibly

preferential voting – may be more effective means to such democratic *desideratum*.

Chapter 4

Who Runs for Higher Office? Electoral Institutions and Level-Hopping Attempts in Germany's State Legislatures

Who Runs for Higher Office? Electoral Institutions and Level-Hopping Attempts in Germany's State Legislatures

Leonardo Carella[†]

Abstract

Do electoral institutions matter for subnational politicians' career choices in a multi-level polity? The paper considers this question by analysing candidacies of sitting German State legislators for the Federal parliament ('level-hopping attempts'), leveraging within-legislature variation in electoral rules due to the widespread adoption of mixed-member systems in Germany's subnational parliaments. State MPs elected via list PR can be expected to be more likely to attempt level-hopping than those elected in the single-member districts (SMD) tier, as the former face lower re-election rates and the latter are more directly accountable to their constituency's voters. Empirical evidence from a novel dataset of State legislators spanning ten Federal elections (1987-2021) confirms this hypothesis. Moreover, the analysis shows that the difference in behaviour across tiers is more marked when State MPs run for insecure Federal candidacies than when they are offered secure candidacies. The findings suggest that subnational electoral institutions play a role in enabling or constraining progressive ambition. Moreover, they highlight a previously overlooked dimension of the 'mandate divide' between MPs belonging to different electoral tiers in mixed-member systems.

[†]Mansfield College and University of Oxford. email: leonardo.carella@mansfield.ox.ac.uk

4.1 Introduction

One of the key insights from legislative studies is that the incentive and opportunity structures that shape legislators' behaviour vary according to the electoral rules that elected representatives are subjected to. Most significantly, electoral institutions determine the *competitiveness* of the reelection and re-selection processes that legislators face and the nature of *accountability* that they are subject to (André, Depauw and Shugart, 2014, p.232). This paper draws on the argument that electoral institutions shape the conduct of elected representatives, and extends the analysis to a previously overlooked aspect of legislators' behaviour: level-hopping attempts. A 'level-hopping attempt' is defined as *the candidacy of a sitting legislator to an elective office at a higher territorial level of politics in a multi-level polity.* In this sense, level-hopping attempts are a form of behaviour that reveals the progressive ambition (Schlesinger, 1966) of an MP: their preference for a career in a 'higher' legislative arena than the one they currently serve in. In short, the paper suggests that there is a connection between electoral institutions of sub-national legislatures and level-hopping attempts: the progressive ambitions of MPs elected from single-member districts are more constrained than those of legislators elected via PR lists.

There are empirical and normative reasons to identify and explain the phenomenon of level-hopping attempts. Empirically, they provide us with an indicator of politicians' willingness to 'move up' on the career ladder, which is distinct from two of the most commonly studied variables in the literature on multi-level careers. On the one hand, level-hopping attempts are analytically distinct from actual career movement across levels, insofar as they comprise instances of careerist behaviour that are frustrated by the circumstances and therefore are not identifiable simply by looking at career trajectory of politicians (Stolz, 2015). On the other hand, a candidacy is a revealed preference, as opposed to measures of stated progressive ambition drawn from legislator surveys (Maestas, 2003). In this sense, the paper contributes to the literature on multi-level careers by analysing a *behavioural* indicator of progressive *ambition.* Moreover, understanding the predictors of level-hopping attempts is arguably also important because such behaviour can be normatively undesirable: legislators who try to forsake their mandate mid-term in pursuit of career advancement are clearly failing to keep their side the election 'bargain' with voters. For instance, Høyland, Hobolt and Hix (2019) find that members of the European Parliament who harbour career ambitions at national level are less likely to attend and participate to legislative work in the body they were elected to.

The phenomenon of level-hopping attempts is investigated in the German context, a case of multi-level polity with a Federal parliament (*Bundestag*) elected via mixed-member system with a closed-lists PR

tier and State legislatures (*Landtage*, singular: *Landtag*) that, in most cases, mirror the Federal electoral rules. Mixed-member systems present an interesting case of within-legislature variation in institutional incentive structures, and the literature has identified systematic differences between legislators elected from the single-member district tier and those elected from the list PR tier with respects to a range of aspects of legislators' behaviour (Lancaster and Patterson, 1990; Stratmann and Baur, 2002; Lundberg, 2006; Maaser and Stratmann, 2018; Breunig, Grossman and Hänni, 2022). In keeping with the thesis' focus on electoral institutions, this paper examines primarily the relationship between tier of election at State level – PR list or single-member district – and level-hopping attempts. By linking sub-national electoral systems with progressively ambitious behaviour, this paper makes two contributions. First, it highlights yet another behavioural difference between list PR and SMD legislators in a relatively rare context of within-legislature variation in electoral rules at sub-national level. Secondly, it centres sub-national electoral systems as an important institutional variable shaping legislators' career incentives and options in a multi-level polity.

The starting point of the argument is that legislators make career decisions not only on the basis of the perceived relative value of the current and prospective post, but also on the perceived probability of retaining those posts in future periods. I further posit that electoral tiers in a mixed-member system differ along two key dimensions related to career prospects: the lower electoral security associated with the list PR tier and the higher degree of personal accountability of a legislator associated with the single-member district tier. As discussed in section 4.4, these two factors should contribute to constrain single-member district MPs' progressive ambitions and, conversely, to encourage list MPs' candidacies to the Federal level. The resulting hypothesis that list PR MPs are more likely candidates for level-hopping attempts comes with a qualification: this difference in progressively ambitious behaviour should be particularly marked for insecure Federal candidacies. High-quality, secure candidacies will be equally appealing to both list and district legislators: they come with the prospect a long career in a more powerful legislative arena than the one they currently operate in, and – by ensuring successful level-hopping – reduce the costs associated with an unsuccessful Federal candidacy.

To conduct the analysis, I compiled a novel dataset of legislators in German State Parliaments with data on over 8,000 MPs spanning almost forty years. This resource, which is perhaps the core contribution of the study, is employed for a two-step empirical investigation into the relationship between electoral tier and legislators' career trajectories. First, the assumption that list PR legislators face lower prospects of re-election at State-level is tested on a sample of legislators sitting in the State parliaments at the end of each term,

providing robust evidence in the expected direction. Secondly, I combine the State Parliament membership data with information on Federal candidacies in ten elections (1987-2021) and State-level leadership positions to test the relationship between tier of election level-hopping attempts. The multivariate analysis supports both the argument that list PR legislators are more likely to run for a Federal seat, and the further hypothesis that the effect of electoral tier is stronger for insecure candidacies than for secure ones.

The paper proceeds as follows. In section 4.2, I contextualise this study within the two lines of research it aims to speak to: the scholarship on progressive ambitions in multi-level politics and the literature on electoral tier effects on legislator behaviour in mixed-member systems. Section 4.3 provides background information on Germany's legislatures, electoral systems and candidate selection practices. In section 4.4, I present the theoretical framework and derive some hypothesis to put to empirical test. Section 4.5 details the data collection project, describes the procedures employed to link State parliament membership data with the other data sources, and outlines the methodological choices of the empirical analysis. Sections 4.6 and 4.7 presents the findings of, respectively, an analysis of reelection rates and level-hopping attempts of German State legislators. The results support the assumption that list PR MPs are less likely to be reelected at State level, as well as the hypotheses that list PR MPs are more likely to run for Federal office and that the effect is more significant for insecure candidacies relative to secure ones. Section 4.8 concludes, discussing implications and limitations of the analysis, as well as avenues for future research.

4.2 Related Literature

As mentioned, this paper speaks to two areas of research: the literature on the electoral tier divides in legislator behaviour in mixed-member systems, and the scholarship on political careers in multi-level polities. With respect to the first line of research, the paper expands the analysis to a new dependent variable: level-hopping attempts. As for the second, this contribution suggests that the electoral institutions of sub-national legislatures are a key variable affecting legislators' tendency to exhibit progressively ambitious behaviour.

4.2.1 Mixed-Member Systems' Electoral Tiers and Legislator Behaviour

A mixed-member system can be defined (Shugart and Wattenberg, 2001, pp. 10-11) as an electoral system where representatives are elected in two overlapping sets of districts (tiers): a nominal tier composed of single-member districts and a list tier composed of multi-member districts. Commonly, the electoral formula of the nominal tier is single-member district plurality, and that of the list tier is closed-list PR.

Originated in post-war Germany (Massicotte, 2003), since the 1990s mixed-member systems – in the two variants of mixed-member proportional and mixed-member majoritarian – have spread to a number of new democracies (Ukraine, Lithuania, Mexico, Russia, Taiwan) as well as established ones (Italy, Japan, New Zealand). The diffusion of this electoral institution has been accompanied by increasing scholarly interest in the opportunities these contexts presenting within-country variation in electoral institutions may offer to learn more about how electoral systems in general affect the behaviour of voters, parties, candidates and legislators (Stratmann and Baur, 2002). For instance, for Moser and Scheiner (2012, p.45), "the combination of very different electoral rules in the same country provides a unique opportunity to study the effects of separate electoral systems as they operate under identical social, political, and economic conditions". There is however no consensus as to whether and under which circumstances the tiers of mixed-member systems can be assumed to approximate the incentive structures of different sets of electoral rules (the 'controlled comparison' thesis), as opposed to producing unique institutional environments via their interaction (what has come to be known as the 'contamination effects' argument) (Herron, Nemoto and Nishikawa, 2018).

Studies on legislators' behaviour in mixed-member system reflect this broader tension in the literature. On the one hand, the 'controlled comparison' approach suggests that differences in how MPs elected in different tiers behave should mirror differences in incentive structures associated with the electoral rules of those tiers. In this perspective, differences in behaviour across tiers should reflect the competing principals that MPs elected from 'pure' list PR and 'pure' single member districts are also primarily accountable to: respectively, their party and their local electorate (Carey, 2007; Batto, 2012). This perspective – which is sometimes known as the 'mandate divide' hypothesis – purports to explain why list PR legislators tend to exhibit higher levels of party discipline (Batto, 2012), to be less responsive to constituents (Breunig, Grossman and Hänni, 2022), and to select committees involved with their party's core issues and constituencies over those that can deliver locally targeted goods (Stratmann and Baur, 2002; Maaser and Stratmann, 2018). These arguments have been criticised on the grounds that the 'mandate divide' approach obscures possible contamination effects: in particular, dual candidacies may create more complex incentive structures for legislators that cannot be flattened onto a PR-SMD distinction. As most mixed-member systems allow candidates to run simultaneously on both tiers, "dual candidacy allows the candidate to hedge bets, particularly if he or she is highly placed on the PR list: the candidate could lose one race yet still gain a seat in parliament. On the other hand, because of dual candidacy, a legislator might then need to satisfy both a national party and a local constituency, to varying degrees" (Herron, 2002, p.367). From this perspective, behavioural divides

associated with mixed-member system tiers tell us little about systematic differences between the effects of PR and SMD systems; rather, they simply reflect the unique institutional configuration of an individual mixed-member system.

4.2.2 Multi-Level Careers and Progressive Ambition

Unlike voting behaviour and committee membership – the most common dependent variables in the 'mandate divide' literature – this paper considers a behavioural variable that has little to do with policy, but rather centres political actors' self-interested goals of professional advancement. The study of politics as a career dates back at least to Max Weber's 1919 *Politics as a Vocation* lecture (Weber, 2008). The theoretical development of this perspective owes much to American behaviouralism (Ruchelman, 1970; Black, 1972; Rohde, 1979), and specifically to Joseph Schlesinger's seminal study of US executive and legislative officials' career pathways in the book *Ambition and Politics*, which starts with the oft-quoted maxim that "ambition lies at the heart of politics" (Schlesinger, 1966, p. 1). In particular, Schlesinger famously distinguished between *discrete*, *static* and *progressive* ambition: the first designates politicians' wish to hold a political job for the duration of the mandate, the second to their desire to retain the office they hold in the future, while the third refers to their aspiration to attain a more important post (Schlesinger, 1966, pp. 9-10). Because of its obvious connection to upward career movement, progressive ambition has been the key concept of interest for research into multi-level careers.

Americanists working in this tradition have found that progressive ambition affects the voting behaviour (Francis and Kenny, 1996), legislative activity intensity (Herrick and Moore, 1993) and district opinion responsiveness (Maestas, 2003) of US representatives aiming for higher office. Less is known about the individual characteristics that drive progressive ambition: however, there is evidence that factors normally associated with 'nascent' political ambition among members of the public, such as gender and personality traits, are not as predictive of progressively ambitious behaviour and attitudes among office-holders (Fulton et al., 2006; Dynes, Hassell and Miles, 2019). Moreover, ambition theory has been influential in the study of the professionalisation of US State legislatures: that, is the extent to which these bodies offer its members opportunities for full-time lifelong employment in the same way Congress does (Squire, 1992). For instance, Schlesinger's discrete-static-progressive ambition typology maps onto Squire's (1988) distinction between *dead-end legislatures*, where professionalisation is low and turnover is high, *career legislatures*, characterised by high professionalisation and low turnover, and *springboard legislatures*, where personnel turnover is high

not because these bodies do not offer satisfying career options but because they commonly serve as stepping stones to higher office.

In the past few decades, the study of multi-level careers has expanded beyond the American context, focussing particularly on the institutional determinants of integration of national and sub-national personnel trajectories (Samuels, 2003; Borchert and Stolz, 2011*c*; Edinger and Jahr, 2016). The recent interest in the territorial dimension of the organisation of careers reflects processes of regionalisation of unitary states, professionalisation of state legislatures in federal ones, and the emergence of supranational political arenas such as the European Parliament (Swenden, 2006; Stolz, 2013; Høyland, Hobolt and Hix, 2019). The extant research has highlighted how institutional factors – federalism, legislative professionalisation, party selection procedures, office accumulation rules etc. – shape the *accessibility*, *availability* and *attractiveness* of political offices, which in turn affect the cost-benefit calculations behind politicians' career choices (Borchert, 2011). Much of this work is limited to single-country case studies; therefore, not much is known in comparative terms with regards to, for instance, the relationship between electoral systems and multi-level career movement. The case studies, however, do suggest that the American case, where career movement across territorial levels is essentially centripetal, is not a common template (Borchert and Stolz, 2011*b*).[1] If anything, what is striking is the cross-national diversity of patterns of elite circulation: alongside the unidirectional model, common to the US and some unitary states prior to devolution (Oñate, 2018; Lo Russo and Verzichelli, 2016), we find contexts where national and sub-national levels function as separate arenas for 'alternative' career progression patterns (Docherty, 2011; Stolz, 2011), or interact in a more complex way, with frequent movements up *and down* the territorial ladder (Santos and Pegurier, 2011; Vanlangenakker, Maddens and Put, 2013; Dodeigne, 2014; Di Capua et al., 2022).

4.3 The German Case

Federalism is one of the core constitutional features of the German polity: its sixteen States (*Länder*, singular: *Land*) enjoy comparatively high levels of autonomy from the Federal government in Berlin in terms of their administrative and policy-making prerogatives (Ladner, Keuffer and Baldersheim, 2016). State Parliaments are a key institution of this level of governance: these are directly elected bodies that pass

[1] One key difference between the American context and most European multi-level polities that makes it hard to generalise from the US literature is that the timings of Federal and State legislative elections in the US are largely synchronised, while elsewhere the political cycles of national and sub-national politics are staggered. US legislators thus have decide *between* elections at the end of one legislative term, while (for instance) for German State MPs running for the Federal parliament is not mutually exclusive with respects to re-running for the State parliament in the future.

legislation in some policy areas (education, policing, planning, local government, culture and church affairs), as well as electing and monitoring State governments, which in turn implement both Federal and State law. Nonetheless, there is a clear hierarchy between the Federal and the State level: "In the German context it is very common to label the move from a State parliament to the national legislature, the Bundestag, as career advancement, whereas a move in the opposite direction is usually considered a step backward" (Jahr, 2015, p.55). Although there is some debate on whether the legislative powers of the Landtag have diminished over time (Reutter, 2006), the consensus is that these bodies have become increasingly professionalised since the 1970s (Borchert and Stolz, 2011*a*). Table 4.1 provides some information on the level of professionalisation of the institutions (MP allowance, number of yearly committee sessions, MP-to-population ratio), as well as detailing the characteristics of their electoral systems. The comparison with the Bundestag shown in the table highlights significant variation between States, which range from Landtage of larger states such as North Rhine Westphalia and Bavaria, where legislator pay and MP-to-population ratio approach those of the Federal Parliament and legislative activity is intense, to the legislatures of the three City-States (Berlin, Bremen and Hamburg), which clearly appear less professionalised.

Thirteen of the 16 Landtage are elected via a mixed-member system with a compensatory PR tier, which makes them a rare case of sub-national parliaments that present within-legislature variation in the electoral rules, alongside the Scottish and Welsh Parliaments in the UK and municipal assemblies of major cities in Hungary and South Africa. The modal Landtag has an electoral system mirroring closely that of the German Federal Parliament: a mixed-member system with around half the legal number of MPs elected in the single-member district tier, and a compensatory closed-list PR tier with a 5% threshold. Voters cast two votes: a first vote (*Erststimme* or *Direktstimme*) for their single-member district candidate, and a second vote (*Zweitstimme* or *Listenstimme*) for a party list: normally a single one for the whole state but in some cases multi-member PR districts within the State have separate lists. There are some notable exceptions to this formula. Saarland and the City-States of Hamburg and Bremen use pure PR formulas. Moreover, the mixed-member systems of Baden-Württemberg and Bavaria apportion seats within party lists in ways that depart significantly from the closed-list formula. In Baden-Württemberg, voters cast only one vote for both a party and for a district candidate: within each party's list, compensatory PR seats are allocated to candidates who failed to win their single-member district but obtained the highest shares of their district vote. Bavaria employs a peculiar open list ballot formula: the second vote also allows voters to express a preferential vote for individual candidates within the list of their choice; PR seats are attributed on the basis

Table 4.1: German State Parliaments (1980-2021).

	Electoral System Characteristics						*Legislature Professionalisation Indicators*		
	Electoral system	**Seats (SMD)**	**PR formula**	**PR seat allocation**	**District in list PR tier**	**Term length in years**	**Seats/100k residents**	**Allowance p.m. (2022)**	**Committee sessions p.a.**
Brandenburg	Mixed	88 (44)	Hare	Closed List	Land	4 → 5 (1994)	3.48	€ 7,604	175
Berlin	Mixed	130 (78)	Hare	Closed List	Land/District	4 → 5 (1999)	3.55	€ 6,657	151
Baden-Württemberg	Mixed	120 (70)	S-Laguë	Best Loser in SMD	Land	4 → 5 (1996)	1.08	€ 7,972	118
Bavaria	Mixed	180 (91)	Hare	Open List	District	4 → 5 (1998)	1.37	€ 8,886	289
Bremen	PR	84 (0)	S-Laguë	Flexible List	District	4	12.35	€ 5,150	179
Hessen	Mixed	110 (55)	Hare	Closed List	Land	4 → 5 (2003)	6.53	€ 8,297	199
Hamburg	PR	121 (71)	S-Laguë	Flexible List	Land/District	4 → 5 (2015)	1.75	€ 3,555	185
Mecklenburg WP	Mixed	71 (36)	Hare	Closed List	Land	4 → 5 (2006)	4.41	€ 6,466	227
Lower Saxony	Mixed	135 (87)	d'Hondt	Closed List	Land	4 → 5 (1998)	1.69	€ 7,175	352
NR Westphalia	Mixed	181 (128)	S-Laguë	Closed List	Land	5	1.01	€ 9,603	322
Rhineland Palatinate	Mixed	101 (51)	S-Laguë	Closed List	Land/District	5	2.46	€ 7,394	148
Schleswig-Holstein	Mixed	69 (35)	S-Laguë	Closed List	Land	4 → 5 (2000)	2.37	€ 8,886	172
Saarland	PR	51 (0)	d'Hondt	Closed List	Land/District	5	5.18	€ 6,238	241
Saxony	Mixed	120 (60)	d'Hondt	Closed List	Land	4 → 5 (1994)	2.96	€ 6,237	125
Saxony Anhalt	Mixed	83 (41)	Hare	Closed List	Land	4 → 5 (2006)	3.81	€ 7,230	205
Thuringia	Mixed	88 (44)	Hare	Closed List	Land	4 → 5 (1994)	4.15	€ 6,036	127
Bundestag	Mixed	598 (299)	S-Laguë	Closed List	Länder	4	0.72	€ 10,012	738

Notes: **(1)** The legal number of seats may not correspond to the number of legislators due to compensation mandates for overhang seats; see Reutter (2021, pp. 37-47). Over the time period considered the number of seats changed for the following parliaments: Berlin, 150 (of which 90 SMDs) until the 1999 election; Bavaria, 204 (104) until 2003; Bremen, 100 (0) until 2003; 83 (0) until 2015; Lower Saxony, 155 (100) until 2003; North-Rhine Westphalia, 201 (151) until 2000; Saxony-Anhalt 99 (49) until 2006 and 91 (45) until 2016; Schleswig-Holstein, 75 (45) until 2000; Germany, 656 (328) until 2002. **(2)** The formula for PR seat allocation changed in Baden-Wurttenberg (d'Hondt until the 2011 election); Bremen (d'Hondt until 2003); Hessen (d'Hondt until 1983); Hamburg (d'Hondt until 2008); Lower-Saxony (Hare until 1982); North-Rhine Westphalia (Hare until 2010); Rhineland Palatinate (d'Hondt until 1991, Hare until 2011); Schleswig-Holstein (d'Hondt until 2009); Germany (d'Hondt until 1987, Hare until 2009). **(3)** The ballot formula changed in Hamburg (closed list until 2011) and Bremen (closed list until 2011). **(4)** Parties can decide whether to field statewide or districtwide lists in Berlin and Rhineland Palatinate; Hamburg and Saarland have two PR tiers (one statewide, one districted). **(5)** The number of committee sessions per annum is an average of the years from 1990 to 2020. *Sources*: Committee sessions per annum from Appeldorn and Fortunato (2021); the rest is the author's own compilation from official sources.

of candidates' preferential votes *plus the candidates' first votes in their district* (for details see Rudolph and Däubler, 2016).[2]

The process of candidate nomination is broadly similar at Federal and State level: the same actors, the regional party branch and the State party, control respectively the selection of district candidates and the drafting of PR lists for both types of election (Detterbeck, 2012, pp. 156-170). The candidate nomination process starts with regional party organisations' selection of district candidates for the seats that fall in their territorial remit through a secret ballot of the local party membership or their delegates.[3] "In rural areas and for a direct seat in the Bundestag this regional party organization will often be identical with the constituency party organization. But in metropolitan areas and for most direct seats in Länder parliaments the regional party organization will supervise selection in several districts" (Patzelt, 2007, p. 52). This process builds a strong linkage between the regional associations and the district MPs after the election: the former actively discourage challenges of incumbents unless there is genuine dissatisfaction with their record; the latter, in turn, will dedicate themselves to constituency work and more generally cater to their selectorates in their legislative activity to secure non-competitive 'coronations' in future selections (Roberts, 1988; Detterbeck, 2016). After this process, a statewide party convention drafts the PR list(s). The details of the selection procedure varies across parties and States, but broadly speaking differs from the district nominations in two respects. First, it is much more top-down than the regional party conventions: "Land party elites are in control of balancing party lists. While delegates decide, they often follow the suggested order of rankings" (Detterbeck, 2016, p. 840). Secondly, while incumbents do enjoy a clear advantage in securing high list positions, State parties use this process to 'rebalance' the outcome of the district candidate selection, taking into account a set of competing formal rules and informal norms in drafting the lists. The most common of these are gender and regional quotas, but parties like the Greens and Die Linke also have rules aimed at guaranteeing winnable list positions to newcomers and younger aspirants (Reiser, 2014). Moreover, at this stage party elites can parachute experts and non-party members in high list positions (Detterbeck, 2016). In sum, obtaining a winnable list position is a more accessible task for outsiders and a less secure prospect for incumbents than securing a safe district nomination.

[2]For the purposes of the analysis, Bavaria – but not Baden-Württemberg – is considered to have a preferential vote system in the PR tier, together with Bremen and Hamburg post-2011.

[3]In Hamburg, Bremen and Saarland, local branches select district lists rather than single-member district candidates.

4.4 Theory

The decision to seek a Federal candidacy can be conceived as a cost-benefit calculation, where legislators weigh the expected utility of continuing their career at State level against the expected utility of a 'candidacy lottery', which yields a Federal position in the event of success and the continuation of their career as State legislators in the event of failure. Two key assumptions must be made at this stage. First, legislators' horizons extend beyond the current period: they care about the value of a political position *and* they care about retaining it in a number of future periods, which is conditional on the expected electoral security of the posts (in Schlesinger's terms, they do not simply have discrete ambitions). Secondly, level-hopping attempts are costly, *and especially so if unsuccessful*. The costs of a Federal candidacy include:

— the *material* costs – effort, time, exposure and financial investment – required to pursue nomination and run in an electoral campaign, which only yields returns if the candidacy is successful.[4]

— the *opportunity* costs (Lazarus, 2006) of pursuing a Federal nomination as opposed to, for instance, using that political capital to achieve posts of influence at State level or advance legislative goals in their current position.

— the *psychological* (Kim, 1970) and *reputational* (Black, 1972) costs of an election loss. Indeed, for State legislators the latter may be compounded by the fact that level-hopping attempts reveal progressive ambition, signalling that they consider their post as a springboard to another legislative body.

4.4.1 Electoral Security

A first possible source of divergence between list PR and SMD legislators' likelihood of attempting level-hopping is the 'baseline' electoral security associated with the two tiers: i.e. legislators' re-election prospects in the State legislature independently of their decision to run for a Federal seat. Clearly, a legislator at risk of exiting the State parliament will be more eager to seize the opportunity of a Federal election to avert such an eventuality than one who can expect a long and secure career in sub-national politics. Empirically, multi-member PR seats have in fact been found to be associated with lower re-election rates than single-member district seats both across countries employing different systems and across electoral tiers in mixed-member systems (Matland and Studlar, 2004; Manow, 2007). As shown in section 4.6, I find this to be the case for

[4] In Germany, candidates are allowed to draw on personal funds and raise funding from private parties. These may account for a large portion of the campaign budget (Zittel and Gschwend, 2008; Van Erkel, Thijssen and Van Aelst, 2017).

German State legislatures as well, where re-election rates of list PR MPs are over 10 percentage points lower than those of SMD legislators.

One explanation for the different levels of electoral security of PR and SMD legislators advanced in the literature concerns SMD MPs' stronger incumbency advantage, which in turn reflects the higher degree of control they have over their re-selection and re-election relative to list PR MPs. As far as re-selection is concerned, the candidate nomination process for single-member districts is significantly less competitive for incumbents, as parties are wary of replacing district MPs that can draw on a personal vote (Heinsohn and Schiefer, 2019): for instance, in German single-member district nominations "there seems to be a strong norm not to challenge incumbents without good reason" (Baumann, Debus and Klingelhöfer, 2017, p. 983). Conversely, as there are more list candidate spots than incumbents, the process of drafting a candidate ranking is likely to present a higher degree of competition for winnable positions, which may result in incumbents being moved down the list. Analogously, SMD incumbents have also more control over their re-election than list PR legislators. As the nominal vote (*Erststimme*) is to some degree a personal vote, SMD MPs can use the resources at their disposal – campaigning, committee assignments, position-taking – to maximise re-election chances. Conversely, list PR seats are attributed on the basis of the 'party' vote (*Zweitstimme*), which reflects voters' partisan preferences and is broadly inelastic with respects to individual MPs' efforts (Zittel and Gschwend, 2008). While the argument for an incumbency advantage of SMD over list PR can be generalised across electoral system families, there is a second explanation as to why list PR MPs may be expected to face lower re-election prospects that is specific to mixed-member systems: the implications of dual candidacies. Because in mixed-member systems district seats are assigned before list PR seats, a 'dual candidate' who would be elected via both tiers ends up taking the district mandate. It follows that district MPs may or may not have secured a 'fallback' post in a safe position on the lists, while list PR legislators by definition have either ran for a district they failed to win or failed to obtain one at all. Thus, on average, sitting SMD legislators can be expected to have secured a more favourable combination of district and list candidacies than sitting list PR MPs, and are therefore likely to retain such advantage in future electoral contests.

4.4.2 Constituency Commitment

Alongside different levels of 'baseline' electoral security, it may also be possible that the nature of accountability engendered by the two tiers affects legislators' decision-making calculations. District MPs owe their

post as candidates to regional party branches and their post in parliament to the 'first vote' of the voters from their district. Both these actors can be assumed to have strong preferences for a locally-oriented legislator. It follows that signalling commitment to the local constituency is crucial for legislators' future prospect of re-selection and re-election, and therefore there is a clear disincentive to reveal their desire to leave their post mid-term. Because most opportunities to level hop require SMD MPs to let go of their constituency – either by running in the State list or by seeking nomination for an open seat elsewhere in the State – Federal candidacies can be readily interpreted as evidence of low commitment to their current constituency. What is more, if their level-hopping attempt is unsuccessful, they can expect to face a competitive re-selection process and a more uphill re-election contest to retain their post at State level. As Vanlangenakker, Maddens and Put (2013, p. 364) put it, "the constituency MPs can build up a strong personal network in their district, thereby increasing their chance on re-election. However, an MP will lose this incumbency advantage if he or she does not stick to the same constituency and thus to the same election." List MPs, conversely, are largely shielded from the reputational costs of a failed Federal candidacy, as the key gatekeeper to their permanence in the legislature is the State party. Unlike district voters and regional branches, which have an interest for a locally oriented legislator, State parties can be assumed to be indifferent to a legislator's desire to 'move up' as she would still be accountable to the same party organisation in the Bundestag as in the Landtag. Moreover, in a closed-list PR system, voters cannot punish progressively ambitious behaviour of an individual legislator with their second-vote. In a way, this *constituency commitment* mechanism is simply the flip-side of the 'incumbency advantage' argument outlined above: SMD legislators may have more control over their re-election prospects, but this also means that they can damage those prospects by acting against the interests of their local (s)electorates. In other words, a Federal candidacy is not only relatively less attractive to a SMD legislator, but also potentially more risky.

4.4.3 Hypotheses

In sum, the two tiers present legislators with distinct electoral security environments: being elected from a SMD comes with higher baseline electoral security but also higher reputational costs to a failed Federal candidacy; being elected via a PR list is a more insecure position, but progressively ambitious behaviour is less costly as they do not have to cater to locally-oriented (s)electorates. These factors should map onto higher expected utility from a Federal candidacy for list PR than SMD legislators. Therefore:

Hypothesis 1 *list PR State MPs are more likely to attempt level-hopping than SMD MPs.*

A further implication of these considerations concerns the *type* of candidacy that legislators from the two tiers can be expected to take up. The two hypothesised sources of differential behaviour between list and district MPs – different reputational costs and baseline electoral security – are only realised if the Federal candidacy is unsuccessful. It follows that, if a State legislator enters the Federal nomination process expecting to secure a Federal seat with high probability, these variables will weigh less on their considerations, as candidacy failure becomes unlikely. Therefore, for 'high-value' candidacies, we should observe smaller differences in the behaviour of legislators from the list PR and district tiers than for 'low-value' candidacies. Long-shot candidacies, on the other hand, will be mostly appealing to list PR legislators, whose permanence in the State legislature is at higher risk if they miss a level-hopping opportunity and who have relatively less to lose from trying to seize the chance when a Federal election comes up.

Hypothesis 2 *list PR State legislators are more likely to attempt level-hopping with an insecure Federal candidacy than SMD legislators, while the two types of MPs are equally likely to run for a secure Federal candidacy.*

4.5 Data and Methods

The empirical section of the paper proceeds in two steps: section 4.6 presents a short analysis of the relationship between State legislators' tier of election and their electoral security at State level, which is a core premise of the theoretical argument; section 4.7 proceeds to test the hypothesised relationships between tier of election, level-hopping attempts and candidacy quality. The first part relies entirely on the novel database of German State legislators, which allows to investigate the likelihood of permanence in sub-national legislatures across 129 State elections. The second section combines Landtag membership data with information from three more datasets, which record (1) Federal candidacies, (2) membership of State executives and (3) parliamentary leadership positions at State level; the analysis encompasses ten federal elections. The data collection and matching procedures employed to generate the samples for the analysis are detailed below.

4.5.1 Datasets

Landtag Membership Dataset

The *Landtag membership dataset* comprises all State legislators who sat in each of the 16 German Länder from the first legislative term starting after the year 1980 until 30 September 2021. It was compiled by

webscraping entries from Wikipedia page directories listing members of German State parliaments for all legislative terms since 1946. Each entry was associated with a link to the legislator's Wikipedia biography, where available, and the text of the webpage was stored. Using the first line of the legislator's Wikipedia biography or – if the page did not exist – a character string of their name and their State, each entry was associated with a legislator ID and a unique legislator-legislature pair ID. I then created a variable recording the number of terms the legislator served in one State parliament, as well as an incumbency status dummy.

For reasons of data availability, further biographical and electoral information was collected only for entries from legislative terms starting after 1980, while data for prior legislatures was set aside. In total, the *Landtag membership dataset* comprises 19,131 legislator-legislature pairs, 8,507 unique legislators[5], and 145 legislatures. Drawing on both the State legislature member directories and legislators' Wikipedia biographies, each legislator-legislature pair was associated with variables recording the party at the time of election and at the end of the legislature, the date of election, the tier and district of election, the dates of permanence in parliament, birth and death dates, and birth place. Exploiting the fact that gender is expressed in German with different articles and word endings, I created a gender variable; moreover, I scraped from Wikipedia biographies alternative or disused personal titles (alternative name spellings, married/unmarried surnames etc.) to facilitate linking across datasets. I integrated this data collection effort with information from official records in case of missing values.

Candidacies Dataset

Official Federal election records compiled by Germany's *Bundeswahlleiter* (Federal Returning Officer) for the ten elections from 1987 to 2021 inclusive were used to compile the *candidacies dataset*. Each entry was associated with list position and single-member district seat, as well as information on the candidacy's outcome (non-election, election via lists, election via SMD). Information on district party shares (*Erststimme*) in each Federal election and in the previous Federal election were associated with entries that competed in a single-member district, using further data from the *Bundeswahlleiter* on real and notional SMD vote shares. After subsetting the dataset to candidates of parties that had representation at State level, candidates were assigned IDs corresponding to State legislator via text string matching.

The State legislator-Federal candidate matching proceeded as follows. In both the *Landtag membership* and *candidacies* datasets, names and surnames – excluding titles (Prof., Dr. etc.) and suffixes (von, van, zu) – were shortened to the first string of the name and the first string of the surname, standardising diacritics

[5]For practical purposes I decided to code legislators who served in multiple State parliaments as different entries.

and other characters that may result in alternative spellings (*ss* for *ß*, *ae* for *ä* etc.). If a candidate and a legislator had the same name, surname and birth year, they were matched. I further recorded whether they matched across any of these combinations of variables: (1) name, surname and party, (2) surname, birth year, party and State, (3) name, birth year, party and State. These 'problem cases' were manually coded as matching or not matching by comparing a range of information available both in the legislators and in candidacies data: married/unmarried surnames, alternative name spellings, occupation, and biographical information on candidacies and Landtag membership in the text of the Wikipedia biography stored in the legislator dataset. Most of matches identified among these 'problem cases' are due to misspellings, alternative spellings, name changes or missing birth year data in the legislator dataset. Out of 27,958 candidacies, 2,975 (10.6%) returned a match with entries in the *Landtag membership dataset*.

Parliamentary and Executive Position Datasets

The *executives* and *parliamentary leadership* datasets list individuals who, in each legislative term and for each Land, held positions of power respectively in the State government and in the State parliaments. The *executives* dataset includes cabinet ministers (*Minister/in*), while the *parliamentary leadership* dataset includes speakers of the Parliament (*Präsident/in*), deputy speakers (*Vizepräsident/in*), leaders of the parliamentary party (*Fraktionsvorsitzende/r*), chief whips (*Geschäftsführer/in*), deputy leaders of the parliamentary party (*Stellvertretende/r Fraktionsvorsitzende/r*), and chairs of standing committees (*Vorsitzende/r des Ausschusses*).

The *executives dataset* was compiled from official sources and Wikipedia directories on State government (*Landesregierung*) composition. The *parliamentary leadership dataset* integrates the one created by Heinsohn and Schiefer (2019), which includes information on those who held parliamentary positions of power in the 1990s and 2000s (corresponding to 4–to–6 legislative terms, depending on the State). Additional research was conducted to cover the period comprised between 1987 and 2021, drawing on data from State parliaments' handbooks, as well as parliamentary and party websites stored on the Wayback Machine internet archive. Ideally, one would wish to know the start and end dates of each post-holder's tenure in office, allowing to isolate post-holders at the time of each Federal election. This information is available for cabinet ministers, but not for parliamentary posts. In this case, both primary (e.g. handbooks) and secondary (the Heinsohn-Schiefer data) sources offer only a snapshot of the allocation of leadership positions at one point in time. Therefore, to maximise the accuracy of the information relative to sitting State legislators *at the time of*

Federal elections, for the *parliamentary leadership dataset*, I collected data on post-holders using sources compiled *prior and as close as possible to* each of the ten Federal elections under consideration. Moreover, I updated some of the Heinsohn-Schiefer data to reflect leadership changes occurring over the course of the parliament. For some legislative terms, however, it was not possible to get such precise information, and the only feasible solution was to retain data on leadership posts collected years prior to the relevant Federal election (provided that this information referred to the State legislative term coinciding with such election).

Entries in the two datasets were assigned IDs corresponding to State legislators using a text string matching technique similar to that outlined above for the matching of election candidates. In this case, post-holders were automatically assigned an ID corresponding to a legislator if they matched across name, surname, party, State and legislative term. Partial matches were dealt with manually with further research. The *executives* dataset contains 1,803 entries (minister-cabinet pairs), corresponding to 903 unique individuals in 134 cabinets: 81.7% of the entries were successfully matched with State legislator IDs. The *parliamentary leadership* dataset contains 4,950 entries (position-legislature pairs), corresponding to a total 2,590 unique individuals over 113 legislatures: 99.2% of the entries were successfully matched with State legislator IDs.[6] The parliamentary posts broke down across positions as follows: 131 speakers (2.6%), 344 deputy speakers (6.9%), 641 party leaders (12.9%), 1,473 deputy leaders (29.6%), 513 chief whips (10.4%) and 1,848 committee chairs (37.3%).

4.5.2 Methods

Electoral Security

The premise that list PR legislators face lower re-election prospects is first investigated descriptively by computing re-election and return rates for list PR and single-member district MPs across subgroups of interest (State, party, gender, seniority, age, term duration). Re-election rates refer to the percentage of MPs sitting in the legislature at the end of legislative term t who are re-elected at $t+1$; return rates refer to the percentage of MPs sitting in the legislature at the end of t who sit in parliament at any point during $t+1$. In essence, the distinction between re-election and return rate is that that the latter includes MPs who narrowly missed election via the list and enter the legislature as substitutes over the course of the term.[7]

[6] The partial matching is mainly due to the fact that Hamburg and Bremen's Parliaments serve both as municipal councils and State legislatures. As the memberships of the two do not fully overlap, some leadership positions may have been held by politicians who are municipal councillors but not State legislators.

[7] The dataset also includes MPs who won their election and did not take up the mandate: these are coded as members from and until the date of the State election, so they are counted as re-elected but not as returned.

Reelection and return rates are clearly an imprecise indicator of aggregate-level electoral security: ideally, we would want to isolate rates of non-re-candidacy and failed re-candidacy. However, because complete State-level candidacy data are not available for most *Länder*, these are the the closest approximation of a measure of electoral security that can be inferred from the data.

Alongside a cross-tabulation of aggregate re-election and return rates, I present the results of logistic models where binary variables for both outcomes (*re-elected* and *returned*), measured at the legislator-legislature pair, are regressed on tier of election and covariates. I introduce controls for State, party and their interaction, and cluster standard errors at the election level. Further Control variables include gender, term duration, election year, preferential vote in the PR tier[8], as well as linear and quadratic operationalisations of age and seniority (number of terms served in the State parliament). I also computed two variables recording the change in party share and number of seats (Δ *Seats*), to account for the fact that incumbent re-election will be more likely when their party gains seats, and control for either of them in alternative specifications of the model. Finally, I specify a model where pure PR elections are excluded, to isolate tier effects in mixed-member systems only.

Level-Hopping Attempts

Section 4.7 presents and discuss the results of empirical tests of the two hypotheses. To reduce the *Landtag membership dataset* to the population of potential level-hoppers, I proceeded as follows. For each of the ten Federal elections considered (1987-2021), I subsetted the data to only those legislators who, for each Federal election, (1) were sitting in State legislatures the day before the Federal election, and (2) belonged to parties that contested the elections. I then merged the ten sets of sitting legislators into single dataset, which therefore has as unit of analysis 'State legislator-Federal election pairs'.[9] The main dependent variable, capturing level-hopping attempts, is a *candidacy* dummy: it takes the value of 1 if a legislator's entry appears among candidates in the relevant Federal election, and 0 otherwise. Using *Bundestag* membership data from the `LegislatoR` repository (Göbel and Munzert, 2022), I recorded whether the level-hopper took up the post in the Federal parliament in a *moved up* dummy variable, which in effect indicates whether the candidacy was successful.

[8]The dummy variable takes the value of 1 for list PR legislators from Bavaria, Hamburg (post-2011) and Bremen (post-2011), who were elected under some type of open or flexible list system.

[9]The unit is distinct from the legislator-legislature pairs used in the *Landtag membership dataset*: a legislator-legislature pair may repeat in the new dataset if the State's legislative term coincides with two Federal election; conversely, legislator-legislature pairs may be dropped if the entire State legislative term falls in between two Federal election, or if an individual legislator had left the State parliament by the date of the Federal election

Moreover, I coded a categorical variable *candidacy quality*, which reflects perceived likelihood of successful level-hopping prior to the election and takes the values of 'Didn't Run', 'Ran in Secure Position' and 'Ran in Insecure Position'. To measure candidacy quality, I examined single-member seat characteristics and seat list position data in the *candidacies dataset*. Single-member seat candidacies are considered secure if the candidate's party won the seat in the previous election (or would have won, under redistricted seat boundaries). State list positions are considered secure if the list position assigned to the candidate is higher than or equal to the last position of the party's elected list candidates in the previous Federal election in a State. To account for dual candidacies across tiers, if either the State list position or the single-member seat are coded as safe, the overall *candidacy quality* takes the value 'Ran in Secure Position'; if neither is safe, the variable takes the value 'Ran in Insecure Position'. To illustrate the coding procedure, consider the example of the CDU members of the North Rhine-Westphalia State Parliament who ran for the Bundestag in the 2002 Federal election in table 4.2. Gerhard Wöchter's candidacy is coded as 'Secure' because in the previous election in 1998 the CDU won the district of Paderborn that he ran in, even though he did not have a position on the State lists in 2002. Laurenz Meyer and Willi Zylajew ran in two constituencies won by the SPD in 1998, so their single-member district races were insecure; however, their candidacies are coded as 'Secure' overall because their position in the list is higher than the position of the last list candidate elected in this tier in 1998 (37^{th} place). Note that this does not mean that the CDU elected 37 MPs from the list tier in 1998 – in fact, the party won 34. It means that, once the constituency MPs received their seat and so could not be elected from the list tier, the 34^{th} list seat was given to the candidate in the 37^{th} position. Finally, Thomas Mahlberg's candidacy is rated as 'Insecure' because his constituency was held by the SPD and his list position in 42^{nd} place would not have guaranteed him a seat in 1998. (And, in fact, it did not win him one in 2002 either.)

Table 4.2: Example of *candidacy quality* coding: level-hopping attempts of CDU members of North Rhine-Westphalia's Landtag running in the 2002 Federal Election.

	Single-Member Seat			*PR List*			*Overall*
Candidate in 2002	District	Winner in 1998	Secure?	Position	Last elected position in 1998	Secure?	Coded as
Gerhard Wöchter	Paderborn	CDU	Yes	None	37	No	Secure
Laurenz Meyer	Hamm - Unna II	SPD	No	4	37	Yes	Secure
Willi Zylajew	Erftkreis I	SPD	No	24	37	Yes	Secure
Thomas Mahlberg	Duisburg I	SPD	No	46	37	No	Insecure

Because likelihood of level-hopping may be related to the utility the legislator assigns to the current

position (moving up to the Bundestag entails giving up the influence they wield in their current job), I also drew on the *executive* and *parliamentary leadership* datasets to code additional variables capturing legislators' power at State level. As discussed, data on position of power at State level have been collected to match as closely as possible the distribution of leadership posts at the time of each election. I coded an *executive position* dummy recording whether the legislator was also a cabinet member at election time, as well as a *party in State government* dummy recording whether the legislator's party was part of the State government majority at the time of election. Using data from the *parliamentary leadership dataset*, I coded two variables recording whether the State legislator held positions of power in the State parliamentary party (*party leadership*, including party leaders, deputy leaders and chief whips) or in the State parliament (*legislative position*, including speakers, deputy speakers and committee chairs). Moreover, I created a *Time to Next State Election* variable recording the expected remaining time a legislator can expect the current State legislative term will last after the Federal election: it stands to reason that the longer a legislator can benefit from their current position, the less attractive a Federal candidacy will appear. The variable, scaled as a decimal fraction of years, is simply computed as the start of the State legislature date plus the legal duration of the State term (4 or 5 years, depending on the State) minus the date of the Federal election associated with each entry.

The empirical section starts by presenting a simple cross-tabulation of level-hopping attempt rates across subgroups, broken down by candidacy quality and candidacy success (table 4.5). I then proceed to test the hypotheses with multivariate regression analysis. To test hypothesis 1, I fit logistic binomial regressions where the probability of running for a Federal seat is modelled as a function of tier of election (list PR vs SMD) and covariates. These include the executive position, legislative position and party leadership dummies, preferential voting rule, gender, time until next State election, as well as linear and quadratic specifications of age and seniority. State, party and Federal election controls are introduced, as well as State $\times$ Party and State $\times$ Federal Election interactions in alternative specifications of the model. Again, I subset the data to mixed-member system legislatures only in one model specification to account for possible differences in the behaviour of PR legislators in mixed-member and pure PR systems. To test hypothesis 2, I ran multinomial regression models where the categorical choice variable *candidacy quality* is regressed on the same independent variables as the binominal model.[10] Heterodaskedasticity-robust standard errors are employed throughout.

[10]In section F of the Appendix, I also re-run this analysis with a categorical variable recording whether the candidacy was successful or unsuccessful. The substantive interpretation of the coefficient for the electoral tier dummy is identical.

4.6 Electoral Security

Table 4.3: Re-election and return rates (sitting legislators at the end of term)

	Re-election Rates		*Return Rates*		*Entries*	
	List PR	SMD	List PR	SMD	List PR	SMD
State						
Brandenburg	50%	62%	55%	65%	302	227
Berlin	46%	66%	50%	67%	810	658
Baden-Württemberg	61%	72%	61%	73%	630	595
Bavaria	62%	75%	63%	76%	818	755
Bremen	60%	—	63%	—	824	—
Hessen	70%	77%	72%	79%	606	501
Hamburg	63%	—	66%	—	1437	—
Mecklenburg WP	50%	63%	55%	65%	233	190
Lower Saxony	55%	68%	59%	70%	554	721
NR Westphalia	49%	66%	57%	67%	670	1055
Rhineland Palatinate	63%	80%	66%	80%	551	255
Schleswig-Holstein	57%	70%	61%	72%	355	353
Saarland	66%	—	70%	—	409	—
Saxony	50%	69%	51%	71%	436	353
Saxony-Anhalt	55%	60%	57%	62%	417	308
Thuringia	52%	71%	58%	73%	289	241
Party						
CDU/CSU	61%	70%	65%	71%	2792	3813
SPD	62%	69%	65%	71%	3565	2114
Greens	53%	76%	56%	76%	1168	70
PDS/Die Linke	60%	66%	63%	69%	734	198
FDP	43%	—	45%	—	723	—
AfD	44%	38%	44%	38%	94	16
DVU/NPD/REP	23%	—	23%	—	115	—
Other Minor Parties	33%	—	33%	—	150	1
Gender						
Female	57%	70%	61%	72%	3089	1117
Male	58%	69%	61%	71%	6252	5095
Seniority						
1^{st} term	56%	76%	60%	78%	4137	1944
2^{nd} term	63%	74%	66%	75%	2185	1601
3^{st} or 4^{th} term	59%	65%	61%	66%	2177	1859
5^{th} term or more	51%	55%	52%	56%	842	808
Age (end of term)						
$<$ 40 years old	63%	82%	67%	84%	1368	431
40−50 years old	67%	82%	71%	84%	2804	1611
50−60 years old	61%	76%	64%	78%	3424	2567
$\geq$ 60 years old	33%	42%	35%	43%	1697	1603
Term Duration						
$<$ 3 years	66%	75%	70%	77%	1066	460
3−4 years	59%	72%	62%	73%	2583	1550
4−5 years	58%	68%	60%	70%	3381	2110
5 + years	53%	68%	57%	69%	2311	2092
Overall	**58**%	**69**%	**61**%	**71**%	**9341**	**6212**

Table 4.4: Logistic Regression Models (clustered s.e. at election level in parentheses)

	Dependent variable:					
	Re-elected			*Returned*		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-20.86^{**}	-21.90^{***}	-29.35^{***}	-26.26^{***}	-27.81^{***}	-34.03^{***}
	(9.13)	(6.89)	(7.08)	(9.25)	(7.23)	(7.27)
List PR	-0.81^{***}	-0.79^{***}	-0.81^{***}	-0.65^{***}	-0.63^{***}	-0.64^{***}
	(0.06)	(0.09)	(0.10)	(0.04)	(0.06)	(0.07)
Seniority	-0.05	-0.08	-0.03	-0.11^{*}	-0.14^{**}	-0.09
	(0.06)	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)
Seniority^2	0.0002	0.003	-0.01	0.005	0.01	-0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age	0.34^{***}	0.34^{***}	0.37^{***}	0.34^{***}	0.34^{***}	0.36^{***}
	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Age^2	-0.004^{***}	-0.004^{***}	-0.004^{***}	-0.004^{***}	-0.004^{***}	-0.004^{***}
	(0.0004)	(0.0004)	(0.0003)	(0.0004)	(0.0004)	(0.0003)
Term Duration	-0.15^{**}	-0.11	-0.03	-0.17^{**}	-0.13^{*}	-0.05
	(0.07)	(0.08)	(0.04)	(0.07)	(0.08)	(0.04)
Gender (Male)	0.06^{*}	0.05^{*}	0.08^{***}	0.06^{*}	0.06^{*}	0.08^{***}
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
Election Year	0.01	0.01^{**}	0.01^{***}	0.01^{**}	0.01^{***}	0.01^{***}
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Preferential Vote	-0.11	-0.17	-0.13	-0.08	-0.14	-0.23
	(0.09)	(0.11)	(0.16)	(0.12)	(0.13)	(0.16)
Δ Seats (share)	5.15^{***}			4.76^{***}		
	(0.46)			(0.47)		
Δ Seats (number)		0.04^{***}	0.04^{***}		0.04^{***}	0.04^{***}
		(0.005)	(0.01)		(0.005)	(0.01)
Nagelkerke pseudo R^2	0.244	0.249	0.264	0.250	0.255	0.269
Observations	15505	15505	12876	15505	15505	12876
State Controls	✓	✓	✓	✓	✓	✓
Party Controls	✓	✓	✓	✓	✓	✓
State × Party Controls	✓	✓	✓	✓	✓	✓
Only Mixed-Member			✓			✓

Note: *p<0.1; **p<0.05; ***p<0.01

A key premise of the theoretical discussion is that legislators' tier of election mark a difference in their re-election prospects, as observed by Manow (2007) for the Bundestag and as it is generally found to be the case in mixed-member systems (Vowles, 2015; Centellas, 2013). The *Landtag membership dataset* allows to examine the plausibility of this assumption for German State parliaments, by comparing re-election and return rates after each State election for the two tiers. Table 4.3 shows the percentage of legislators sitting in the legislature at the end of a parliamentary term who (1) were re-elected in the following legislature (reelection rate) and (2) took up a seat over the course of the next legislature (return rate), grouped by tier of election for some subgroups of interest. Consistently with Heinsohn (2014), patterns of institutional variation suggest that more professionalised State legislatures and shorter effective term duration correspond to lower levels of personnel turnover. Overall, re-election and return rates of SMD legislators are respectively 11 and 10 percentage points higher than those of list PR MPs. Across states, parties, genders, seniority, age and term duration subgroups, for virtually all cases where the sample sizes are meaningfully large, being elected in a single-member district appears more likely to guarantee reelection and permanence in the legislature than being elected via the lists.

The multivariate regression results in table 4.4 essentially confirms this conclusion. Across models specifications of the sample (including or excluding pure PR legislatures) and of the party seat change controls, the log odds coefficients for tier of election are negative for both dependent variables. The average marginal effects for models 3 and 6 are respectively -0.15 and -0.12, implying that holding everything else constant list PR legislators are 15 percentage points less likely to be re-elected and 12 percentage points less likely to be returned to Parliament than single-member district MPs. Within the limitations of the measurement and of the modelling strategy, the estimates are highly significant. As expected, likelihood of permanence in the State legislature is negatively related to term duration and positively related to party change in number or share of seats.

4.7 Level-Hopping Attempts

Let us now turn to a descriptive analysis of State legislators' candidacies to the Federal level from the information contained in the merged dataset. In total, 717 level-hopping attempts were identified – a number which confirms that the phenomenon of level-hopping is relatively rare, but not insignificant. On average, 4% of sitting State legislators run in any given Federal election, and 3.1% of major parties' candidates in each Federal election are sitting State legislators (though with significant variation across parties and elections,

as shown in figure 4.1). More importantly, State legislators tend to win seats relative to the average non-incumbent candidate. Level-hoppers make up a considerable share of each new cohort of members of the Bundestag: on average 11% of newly elected Federal legislators are legislators moving straight from the State parliaments over the period considered. Of the 717 level-hopping attempts, 511 sitting State legislators ran for Federal office only once, 72 twice, 18 three times and two did so four times over the period considered: former Minister-President of Saarland Oskar Lafontaine (SPD, then Die Linke) and CDU member of the Hamburg Parliament Klaus Peter Hesse.

After excluding entries from legislatures for which either data on legislative posts[11] or information on candidacy quality[12] were unavailable, the core dataset used for the rest of the analysis comprises 14,318 legislator-Federal election pairs, with 6,429 unique legislators from 106 distinct State legislatures. Of the 626 level-hopping attempts remaining, 31.2% (195) were in secure positions and 68.8% (431) in insecure positions; 12.7% (55) of candidates in insecure positions effectively moved up to the Bundestag, against 82.1% (160) of candidates in secure positions. Table 4.5 shows how entries, candidacies, candidacy quality and candidacy success break down by electoral tier, State, party, gender, government status, office held, seniority and age groups.

Table 4.5 shows patterns of variation in the frequency of level-hopping attempts as a share of sitting members of State parliaments, as well as in their breakdown across candidacy type. Consistently with the main hypotheses, list PR legislators are over twice as likely to run for Federal seats than single-member district ones, and their candidacies tend to be of lower quality. Level-hopping attempts are also particularly frequent among legislators from smaller states (Berlin, Hamburg, Bremen) and from opposition parties at the State level, which chimes with the intuition that less valuable posts are associated with stronger temptations for level-hopping. MPs from smaller parties (Greens, PDS/Die Linke, AfD, FDP and – particularly – the smaller radical right parties) are more likely to run than CDU/CSU and SPD members of State parliaments. An interesting pattern emerges from the breakdown of candidacies by type of office held by members of parliament in addition to their legislative seat: those with an executive post (cabinet ministers) are somewhat less likely to run than simple backbenchers, while those with a post of influence in the parliamentary party (leaders and deputy leaders) do so more frequently. However, members of the State executive running for Federal office are especially likely to do so in secure positions, while leaders in the party have only slightly

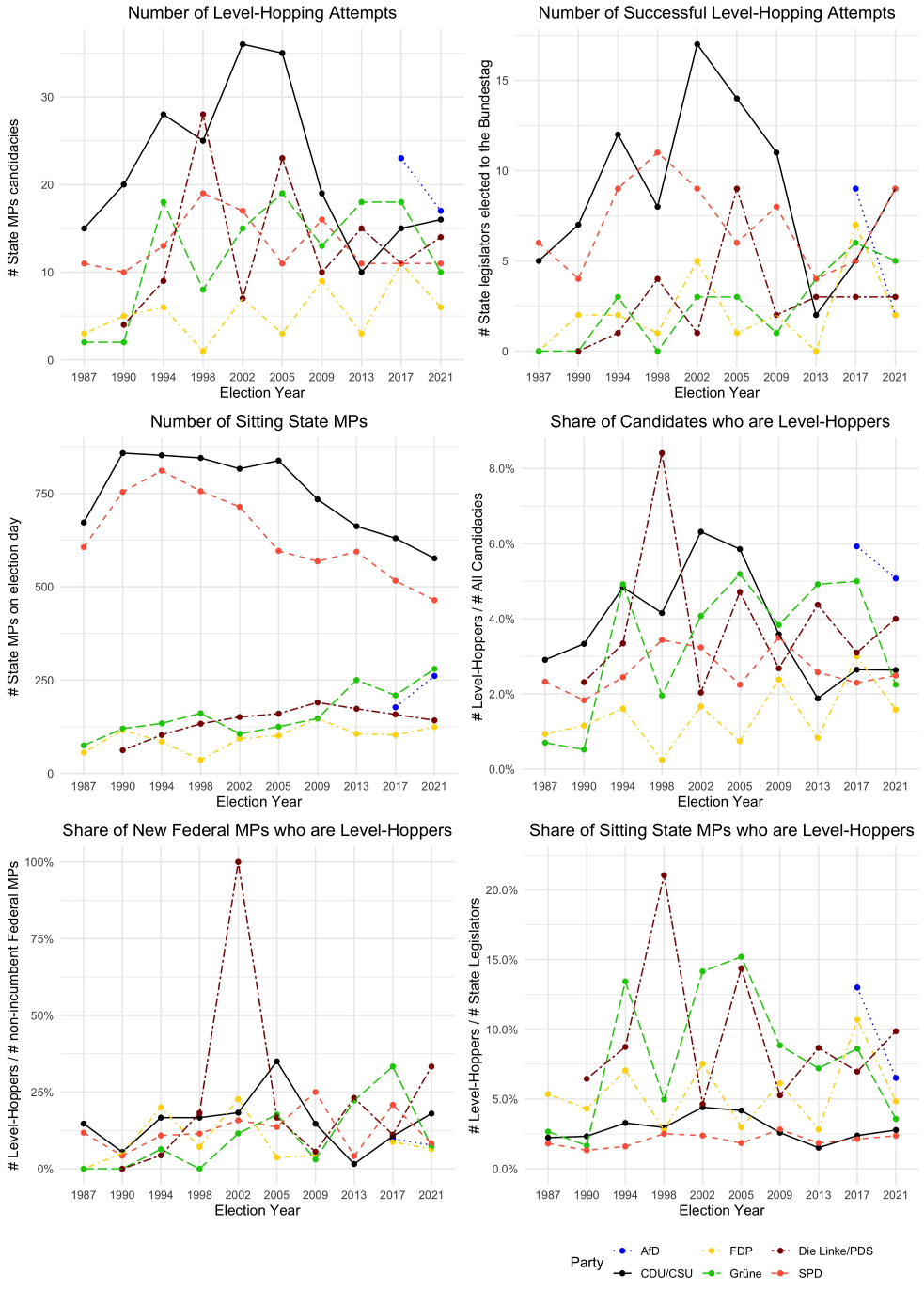
[11] These include State legislatures elected just before the Federal election, where committee assignments had not been realised – it is the case for example of the 17th Bavarian legislature, which was elected just one week before the 2013 Federal election.

[12] These include, for instance, legislators from Eastern States sitting in the first session of the State parliament following reunification in 1990, for which *candidacy quality* could obviously not be coded due to the absence of information on party performance in the previous election.

Table 4.5: Descriptive Statistics: Level-Hopping

	# Entries	Candidacies	% Insecure	% Secure	% Moved up
Electoral Tier					
SMD	6258	131 (2.1%)	42.0%	58.0%	54.2%
List PR (all states)	8060	495 (6.1%)	76.0%	24.0%	29.1%
List PR (Mixed only)	6134	327 (5.3%)	71.9%	28.1%	33.3%
State					
Brandenburg	606	28 (4.6%)	71.4%	28.6%	39.3%
Berlin	1107	113 (10.2%)	72.6%	27.4%	19.5%
Baden-Württemberg	1132	17 (1.5%)	52.9%	47.1%	47.1%
Bavaria	1343	19 (1.4%)	42.1%	57.9%	47.4%
Bremen	680	63 (9.3%)	85.7%	14.3%	15.9%
Hessen	1143	39 (3.4%)	56.4%	43.6%	46.2%
Hamburg	939	95 (10.1%)	85.3%	14.7%	20.0%
Mecklenburg WP	558	31 (5.6%)	83.9%	16.1%	25.8%
Lower Saxony	1220	27 (2.2%)	29.6%	70.4%	81.5%
NR Westphalia	1700	39 (2.3%)	38.5%	61.5%	71.8%
Rhineland Palatinate	808	24 (3.0%)	62.5%	37.5%	45.8%
Schleswig-Holstein	583	24 (4.1%)	62.5%	37.5%	41.7%
Saarland	307	10 (3.3%)	60.0%	40.0%	60.0%
Saxony	743	51 (6.9%)	72.5%	27.5%	35.3%
Saxony Anhalt	829	14 (1.7%)	64.3%	35.7%	42.9%
Thuringia	620	32 (5.2%)	75.0%	25%	28.1%
Party					
CDU/CSU	5629	178 (3.2%)	55.1%	44.9%	42.7%
SPD	4840	109 (2.3%)	45.0%	55.0%	53.2%
Greens	1371	116 (8.5%)	84.5%	15.5%	21.6%
PDS/Die Linke	1081	110 (10.2%)	83.6%	16.4%	21.8%
FDP	766	46 (6%)	69.6%	30.4%	45.7%
AfD	438	40 (9.1%)	87.5%	12.5%	27.5%
DVU/NPD/REP	77	24 (31.2%)	100%	0%	0%
Other Minor Party	116	3 (2.6%)	100%	0%	0%
Gender					
Female	4354	221 (5.1%)	75.1%	24.9%	32.1%
Male	9963	405 (4.1%)	65.4%	34.6%	35.6%
Government Status					
In Government	8246	250 (3%)	59.2%	40.8%	39.2%
In Opposition	6072	376 (6.2%)	75.3%	24.7%	31.1%
Office					
Executive Position	647	18 (2.8%)	11.1%	88.9%	72.2%
State Party Leadership	2282	163 (7.1%)	66.3%	33.7%	36.2%
Legislative Position	3132	165 (5.3%)	66.7%	33.3%	36.4%
None	9693	377 (3.9%)	72.1%	27.9%	31.6%
Seniority					
1^{st} term	5373	230 (4.3%)	81.7%	18.3%	22.6%
2^{nd} term	3746	169 (4.5%)	75.1%	24.9%	30.8%
3^{rd} or 4^{th} term	3779	171 (4.5%)	58.5%	41.5%	44.4%
5^{th} term or more	1420	56 (3.9%)	28.6%	71.4%	62.5%
Age					
$<$ 40 years old	2112	123 (5.8%)	86.2%	13.8%	17.9%
40-50 years old	4038	218 (5.4%)	68.3%	31.7%	36.2%
50-60 years old	5567	223 (4%)	59.6%	40.4%	43.5%
$\geq$ 60 years old	2589	62 (2.4%)	69.4%	30.6%	27.4%
Overall	**14318**	**626 (4.4%)**	**68.8%**	**31.2%**	**39.1%**

Figure 4.1: Descriptive Statistics: election trends in level-hopping by party.



above-average levels of candidacy security. The observed frequency of level-hopping attempts decreases with age and – to a lesser extent – with seniority, though the security of the candidacies these legislators obtain is generally increasing in the same variables.

Table 4.6 presents the regression output for a series of binomial logit models aimed at gauging the change in probability of attempting level-hopping associated with legislators' tier of election. The coefficient for List PR tier of election is positive and strongly significant across alternative specifications of the fixed-effect variables and of the sample, with average marginal effects ranging from a 1.1 percentage point increase (model 5) to a 1.8 percentage point increase (model 3), which are substantial for an event that interests only 4.4% of the observations. These results provide evidence in favour of hypothesis 1. Being in opposition at State level remains predictive of level-hopping attempts net of other variables, as observed in the table 4.5. Positions of power in the executive, state party and the legislature are not consistently associated with the outcome variable. This suggests that holding a position of power at State level may be to some extent counterbalanced by the fact that politicians with ties to the leadership are better positioned to obtain high-quality candidacies. As expected, legislators are more likely to run for a Federal election the closer it falls relative to the next scheduled State election. No significant effects of preferential voting rules or gender emerge from the analysis. Likelihood of level-hopping attempt peaks at 40 years of age, and increases with with the number of legislative terms served in the State Parliament roughly linearly for most of the values of the seniority variable observed.

Hypothesis 2 is tested by substituting a categorical variable taking the values 'Didn't Run', 'Ran in Secure Position' and 'Ran in Insecure Position' to the binary outcome variable in the models 1 and 4 of table 4.6.[13] The regression coefficients in the multinomial model of table 4.7 show therefore the change associated with each independent variable in the log-odds of running in a Secure or Insecure position relative to the referent level 'Didn't Run'. Although subdividing what is already a rare event into categories with quite low cell count comes at the cost of statistical power, in both model specifications, list PR tier increases significantly the probability of running for an insecure position. This does not seem to be the case for the probability of running for a secure position, although the point estimates are also positive. In the full sample model, list PR increases the average probability of an insecure candidacy by 1.8 percentage points and of a secure candidacy by 0.4 percentage points; in the mixed-member system only model, the average marginal effects are respectively 1.12 and 0.37 percentage points. The difference between the two estimates is statistically significant at the 95% confidence level for the full-sample model, but not for the mixed-member only model.

[13]State, Party and Election interactions could not be added due to issues of collinearity.

Table 4.6: Logistic Regression Models (heteroskedasticity-robust s.e. in parentheses)

	Dependent variable:		
	Candidacy (1 = Ran for Federal Election)		
	(1)	(2)	(3)
Constant	-4.35^{***} (1.13)	-3.45^{***} (1.26)	-5.23^{***} (1.37)
List PR	0.48^{***} (0.13)	0.35^{**} (0.14)	0.52^{***} (0.13)
Executive Position	0.30 (0.26)	0.33 (0.27)	0.30 (0.26)
Legislative Position	0.01 (0.11)	0.01 (0.11)	−0.01 (0.11)
State Party Leadership	0.23^{**} (0.11)	0.23^{**} (0.12)	0.23^{**} (0.11)
Party in State Govt.	-0.30^{***} (0.10)	-0.23^{**} (0.11)	-0.26^{**} (0.11)
Age	0.12^{***} (0.04)	0.12^{***} (0.04)	0.12^{***} (0.04)
Age^2	-0.001^{***} (0.0004)	-0.002^{***} (0.0004)	-0.001^{***} (0.0004)
Seniority	0.38^{***} (0.09)	0.38^{***} (0.09)	0.41^{***} (0.09)
Seniority^2	-0.03^{**} (0.01)	-0.03^{**} (0.01)	-0.03^{***} (0.01)
Preferential Vote	0.14 (0.19)	0.16 (0.22)	0.15 (0.20)
Time to Next State Election	-0.15^{***} (0.04)	-0.14^{***} (0.04)	-0.14^{***} (0.04)
Male	−0.06 (0.10)	−0.07 (0.10)	−0.05 (0.10)
Nagelkerke pseudo R^2	0.157	0.183	0.174
Observations	14305	14305	14305
State, Party, Election Controls	✓	✓	✓
State × Party Controls		✓	
Party × Election Controls			✓
Only Mixed-Member			
	(4)	(5)	(6)
Constant	-5.46^{***} (1.40)	-4.79^{***} (1.52)	-6.80^{***} (1.64)
List PR	0.41^{***} (0.13)	0.34^{**} (0.15)	0.47^{***} (0.14)
Executive Position	0.19 (0.28)	0.22 (0.29)	0.16 (0.28)
Legislative Position	0.003 (0.13)	−0.005 (0.13)	−0.02 (0.13)
State Party Leadership	0.11 (0.13)	0.11 (0.14)	0.10 (0.13)
Party in State Govt.	-0.42^{***} (0.12)	-0.37^{***} (0.14)	-0.31^{**} (0.13)
Age	0.16^{***} (0.05)	0.18^{***} (0.05)	0.17^{***} (0.05)
Age^2	-0.002^{***} (0.001)	-0.002^{***} (0.001)	-0.002^{***} (0.001)
Seniority	0.37^{***} (0.12)	0.32^{***} (0.12)	0.40^{***} (0.12)
Seniority^2	−0.02 (0.02)	−0.01 (0.02)	−0.02 (0.02)
Preferential Vote	−0.03 (0.52)	0.37 (0.75)	0.05 (0.55)
Time to Next State Election	-0.17^{***} (0.04)	-0.16^{***} (0.04)	-0.15^{***} (0.04)
Male	0.05 (0.11)	0.05 (0.12)	0.05 (0.11)
Nagelkerke pseudo R^2	0.160	0.187	0.182
Observations	12387	12387	12387
State, Party, Election Controls	✓	✓	✓
State × Party Controls		✓	
Party × Election Controls			✓
Only Mixed-Member	✓	✓	✓
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4.7: Multinomial Regression Models (heteroskedasticity-robust s.e. in parentheses)

	Dependent variable:			
	Candidacy Type (Ref.: *Didn't Run*)			
	(1)		(2)	
	Insecure	*Secure*	*Insecure*	*Secure*
Constant	-3.01^{**}	-14.21^{***}	-3.95^{***}	-12.80^{***}
	(1.22)	(3.29)	(1.51)	(3.45)
List PR	0.80^{***}	0.31	0.60^{***}	0.28
	(0.18)	(0.19)	(0.19)	(0.19)
Executive Position	-1.04	0.54^{*}	-0.89	0.46
	(0.72)	(0.30)	(0.73)	(0.32)
Legislative Position	0.01	-0.03	0.02	-0.05
	(0.13)	(0.18)	(0.16)	(0.20)
State Party Leadership	0.13	0.42^{**}	-0.08	0.38^{*}
	(0.14)	(0.18)	(0.17)	(0.20)
Party in State Govt.	-0.33^{***}	-0.15	-0.48^{***}	-0.23
	(0.12)	(0.16)	(0.16)	(0.18)
Age	0.07^{*}	0.37^{***}	0.11^{*}	0.33^{***}
	(0.04)	(0.11)	(0.05)	(0.12)
Age^2	-0.001^{**}	-0.004^{***}	-0.001^{**}	-0.004^{***}
	(0.0005)	(0.001)	(0.001)	(0.001)
Seniority	0.41^{**}	0.59^{***}	0.63^{**}	0.52^{***}
	(0.17)	(0.15)	(0.28)	(0.17)
Seniority^2	-0.06^{**}	-0.03	-0.09^{*}	-0.01
	(0.03)	(0.02)	(0.05)	(0.02)
Preferential Vote	0.08	0.46	0.40	0.22
	(0.23)	(0.38)	(1.08)	(0.65)
Time to Next State Election	-0.08^{*}	-0.24^{***}	-0.10^{*}	-0.25^{***}
	(0.05)	(0.06)	(0.05)	(0.06)
Male	-0.11	0.04	0.11	-0.07
	(0.12)	(0.17)	(0.14)	(0.18)
Observations	14305		12387	
McFadden's R^2	0.163		0.172	
State, Election, Party Controls	✓		✓	
Only Mixed-Member			✓	
Note:			*p<0.1; **p<0.05; ***p<0.01	

Hypothesis 2 is therefore supported, although the evidence for it is somewhat weaker than for hypothesis 1: list PR legislators are clearly more likely than SMD MPs to run for the Bundestag in insecure candidacy positions; the effect of list PR tier of election on likelihood to run in secure position is – as expected – small and non-significant, but still positive. It is also interesting to note how leadership roles predict differently level-hopping across candidacy type. Party leadership and executive office are positive predictors of a secure candidacy, but the effect on the probability of an insecure candidacy is inconsistent in its direction and non-significant. This suggests that, while overall ministers and members of the State party leadership are neither more nor less likely to run for Federal office than other legislators (table 4.6), *when they do* they are more likely to obtain secure positions. Whether we consider age or seniority, probability of level-hopping attempts peaks later on in a legislator's career with respects to secure candidacies than insecure candidacies.

4.8 Conclusion

The paper presents observational evidence for a relationship between electoral institutions and a previously overlooked aspect of legislator behaviour: level-hopping attempts. The analysis conducted on a novel dataset of members of German State legislatures – where we can leverage within-legislature variation in electoral rules at sub-national level – suggests that, although overall relatively rare in the case under study, these occurrences are more frequent among list PR than SMD legislators. The finding is consistent with the theoretical argument that list PR mandates make legislators' position both more insecure and less directly accountable to voters. Additionally, there is some evidence that the divide in behaviour across electoral tiers is somewhat more pronounced when legislators obtain insecure candidacies than when they are guaranteed upwards career progression with a safe electoral position. Upward movement of political personnel seems therefore in part driven by a logic of necessity and survival: legislators facing electoral risks must seize a level-hopping opportunity when it presents itself, whereas for more established legislators it is often not worth the risk of forsaking a secure sub-national office. Moreover, because under mixed-member rules SMDs are attributed unequally across parties – with the largest political forces getting the bulk of them – a further implication of the findings is that this constituency structure incentivises level-hopping attempts differently across parties. Progressive ambition should thus be more constrained for larger than for smaller parties, and – within major parties – more relevant in States where they are stronger than in those where they are weaker.

As a first investigation into the consequences of sub-national electoral institutions for progressively am-

bitious behaviour of legislators, this study suffers from two limitations. First, the research design does not distinguish whether the relationship between electoral tier and legislator behaviour is due to selection or incentive effects: it might be the case that politicians who take up a SMD seat are less attracted to a Federal career than those who enter the Landtag via the lists to begin with. The results of the analysis of 'secure' and 'insecure' candidacies partly assuage these concerns: when offered secure candidacies, the difference in behaviour between the two types of MPs is smaller and statistically insignificant. This suggests that the more they discount the prospect of retaining their State seat in the future – whichever type it may be – the more the likelihood of progressively ambitious behaviour converges. However, it remains impossible to fully disentangle whether the different electoral tiers attract different types of ambitions or whether they provide legislators that are (on average) equally ambitious with different incentive structures to behave on those ambitions. This shortcoming may be addressed in the future through the collection of complete data on candidacies in State election. These may be used to compare level-hopping behaviour of MPs who narrowly won an SMD seat and those who were narrowly defeated but were nonetheless returned to the Landtag via the PR list tier. However, due to German privacy laws, for most States we can only know the district of MPs elected via the districts and the list position of MPs elected via the lists: candidacies that did not result in an election are expunged from electoral records.

A second limitation due to the unavailability of complete candidacy data for State elections is that it is impossible to isolate dual candidacies. This makes it hard to adjudicate how much of the relationship between tier of election and level-hopping behaviour is due to 'constituency commitment' or 'electoral security' mechanisms. As mentioned, we cannot know from publicly released data whether an SMD legislator held a list position (and how safe such position was), or if a list PR legislator also ran for a single-member district in the previous State election. Further data collection from primary sources may allow, in future research, to address this issue for a sufficient number of State legislatures. Once dual candidacy data is available, it will be possible derive a continuous 'electoral security' variable as a function of probability of re-election in the district and probability of re-election via the lists, along the lines of Stoffel's (2014) 'unified scale of electoral incentives'. If, controlling for such an index, there were still a difference in level-hopping probability across tiers, this could be plausibly attributed to a real divide in how list PR and SMD legislators interpret their mandate; otherwise, one might conclude that level-hopping attempts are a common response of all 'types' of MP to electoral insecurity, and list PR legislators just happen to be more at risk.

Nonetheless, the paper's findings advance our understanding of the role electoral institutions play be-

yond and below the commonly studied level of national politics. To the extent that the findings may be generalised beyond the institutional environment of mixed-member systems, the implication of the argument presented is that sub-national electoral institutions constrain or enable progressive ambition by varying the competitiveness and accountability of an electoral mandate. Institutions like single-member districts, which grant legislators a strong incumbency advantage as long as they display constituency-oriented behaviour, will therefore reduce MPs' temptation to give up their mandate for careerist goals. Institutional settings that attenuate MPs' control over their career progression chances, as we observe in closed-list PR systems, on the contrary will enable progressive ambitions. Sub-national electoral institutions that entrench incumbents may therefore be preferred by constitutional designers who wish to professionalise legislative careers in lower-level parliaments when more attractive options exist higher up the hierarchy of offices of a multi-level polity.

Chapter 5

Predicting Candidate Preference Shares in Preferential-List PR Systems: A Logical Model of Intra-Party Competition

Predicting Candidate Preference Shares in Preferential-List PR Systems: A Logical Model of Intra-Party Competition

Leonardo Carella[†]

Abstract

This paper derives and tests quantitative predictions for three indicators of intra-party competition under preferential-list proportional representation (PLPR) rules: the share of preference votes of the first-ranked candidate (v_1), the effective number of candidates (N_c), and the share of preference votes of the last eligible candidate (v_s). First, it presents a model where these variables are functions of input quantities measurable at election time: the number of candidates c, the number of seats won s and the maximum number of preference votes p. Then, it is shown that these equations can be recast in terms of purely institutional quantities to predict average expected values of v_1, N_c and v_s for any seat-winning list in a district. Given a district magnitude M, a maximum number of preference votes allowed p and a parameter r which captures the permissiveness of over-nomination rules, v_1 is predicted to be $(prM^{\frac{11}{8}})^{-\frac{1}{4}}$, N_c is predicted to be $(prM^{\frac{11}{8}})^{\frac{3}{8}}$ and v_s is predicted to be $(prM^{\frac{5}{2}})^{-\frac{1}{4}}$. These relationships are tested on a diverse sample of data from 31 PLPR elections in nine countries. The model's predictions come remarkably close to describe the empirical relationship between the observations in the sample, and are substantially less biased than those of existing models of intra-party competition. It is also shown that the district-level model can predict *average* values of intra-party competition in a district about as well as the seat-product model predicts analogous inter-party quantities. Implications for further research on the intra-party dimension of electoral systems and institutional design are discussed in the conclusion.

[†]Mansfield College and University of Oxford. email: `leonardo.carella@mansfield.ox.ac.uk`

5.1 Introduction

The seat-product model (SPM) represents a major achievement of the 'Duvergerian agenda', the scholarly quest to identify regularities in the relationship between institutional features of electoral systems and political outcomes (Taagepera and Grofman, 1985; Taagepera and Shugart, 1993; Taagepera, 2007; Taagepera and Sikk, 2010; Sikk and Taagepera, 2014; Li and Shugart, 2016; Shugart and Taagepera, 2017). Through their career's work, Rein Taagepera and Matthew Søberg Shugart – occasionally in collaboration with other authors – have shown that key features of party systems (and the democratic political process more broadly) can be derived deductively from a small set of quantities that characterise an electoral system: primarily, district magnitude (M) and assembly size (S). These relationships between institutional input variables and predicted average outcomes 'in expectation' are normally in the form of $Y = X^k$, an exponential functional form more common to the laws of natural sciences than the linear expressions used in much of social science modelling (Taagepera, 2008, pp. 52-70). For instance, the formula for the expected effective number of parties N_S in an assembly of size S elected from districts of mean magnitude M is $N_S = (MS)^{\frac{1}{6}}$ and for the expected fractional share of the first party σ_1 it is $\sigma_1 = (MS)^{-\frac{1}{8}}$ (Taagepera, 2007; Shugart and Taagepera, 2017).[1] The accuracy of these equations in predicting empirical distributions of party system quantities across repeated elections in large samples is not only substantively interesting for the comparative study of electoral institutions, insofar as they offer a guide as to the expected effects of electoral system reform, but also a testimony to the potential of the theory-building method behind their derivation: logical modelling.

This paper follows in this line of theoretical reasoning, investigating whether features of intra-party competition follow predictable patterns similar to those identified by the seat-product model for inter-party competition. Specifically, it proposes a quantitatively predictive logical model of intra-party competition in preferential-list proportional representation (PLPR) systems, the most common category of preferential voting system. PLPR is defined as an electoral system with the following characteristics: (1) voters can or must cast a personal vote for a candidate within a list of co-partisans, (2) the number of candidates elected in each list is determined by the *pooled* number votes cast at list level, and (3) the attribution of seats to candidates within a list is determined, at least in part, by personal votes (Karvonen, 2004; Shugart, 2005; Passarelli, 2020).

Three quantities of interest that characterise the intra-party distribution of preference votes for a (seat-

[1] I use σ_1 to notate the fractional share of the largest party, whereas this is generally indicated as s_1, to avoid confusion with s (the 'raw' number of seats won by a list, which is central to the model developed in the rest of the paper) and S (the assembly size, which is a key component of the seat-product model).

winning) list competing under preferential-list rules are examined:

— *The share of preference votes obtained by the first candidate in a list (v_1)*: this is simply the number of votes of the first candidate over the total number of preferences cast for the party in that district. The lower this value, the more competitive the list.

— *The effective number of candidates (N_c)*: this is the intra-party analogue of the inter-party notion of 'effective number of parties' (Laakso and Taagepera, 1979), a measure of competition where the count of parties is weighted by their fractional share.[2] In a list with c candidates, N_c is $\frac{1}{\sum_1^c({v_1}^2+{v_2}^2+\ldots{v_c}^2)}$, i.e. the inverse of the sum of squares of candidate shares. The higher this value, the more competitive the list.

— *The share of preference votes obtained by the last eligible candidate (v_s)*: for a list winning s seats in a district, it is the share of preferences gained by the s^{th} candidate.[3] While a higher value of v_s is not *prima facie* an indicator of lower (or higher) competitiveness, the quantity may be of substantive and practical interests for two reasons: in pure open-list PR (OLPR), where only preference votes matter to candidates' election, it identifies the minimum share a candidate must get to win a seat in expectation; in flexible-list PR (FLPR), where threshold constraints apply, it indicates how 'low' the legal threshold should be for the list to function as fully open.

The paper proceeds as follows. In section 5.2, I illustrate the derivation of the basic building block of the existing models of intra-party competition quantities (the Shugart-Bergman-Watt model): the formula for the expected value of first-ranked candidate's preference share (v_1). I then proceed to propose a refinement of this simple model that extends its scope conditions to systems where voters may express more than one preference vote and incorporates an intuitive, but previously overlooked, assumption about the relationship between intra-party competitiveness of a list and actors' expectations of the list's inter-party performance. In section 5.3, I argue that, from v_1, it is possible to compute two additional quantities of interest: the effective number of candidates (N_c) and the preference shares for the last eligible candidate in a list (v_s). Tractable approximations of the equations for these quantities are provided. The resulting models provide 'predictions' specific to each seat-winning list, but rely on two input quantities that are realised at election time – the number of candidates in the list c and the number of seats won by the list s – and therefore are not, strictly

[2] To my knowledge, the first to compute effective number of candidates in a PLPR system was Arter (2013).

[3] I refer to this quantity in terms of 'last eligible' rather than 'last elected' because the model aims to apply to all types of PLPR, including flexible-list systems. In this sub-type of PLPR, there is a threshold of preference votes that candidates must meet to be elected on their preference votes, otherwise the allocation of seats defaults to the candidates' position on the list.

speaking, *pre*-dictive. To address this shortcoming, I proceed to show that these equations can be recast in terms of purely institutional input variables, which vary at district level. I therefore derive two separate models for each of the three quantities of interest (v_1, N_c and v_s): (1) a list-level 'post-result' equation, which returns different values for each list in the same district, and (2) a district-level 'results-blind' prediction, which relies uniquely on institutional variables and predicts the value of the quantity in expectation for *any* list in a district. These sets of predictive equations are tested empirically on a sample of open- and flexible-list electoral outcomes, for a total of over 2,600 seat-winning lists. Section 5.4 describes the data and the institutional characteristics of the countries in the sample. Section 5.5 outlines the empirical modelling choices, and motivates the indicators chosen to compare and assess model fits. Section 5.6 presents the results; the performance of the models' estimates and predictions is compared, in turn, with that of existing models of intra-party competition and with that of the more established predictions of inter-party quantities of the seat-product model. I conclude in section 5.7 highlighting implications and limitations of the analysis.

5.2 Modelling First-ranked Candidate Preference Shares

5.2.1 The Shugart-Bergman-Watt (SBW) Model

Applications of the logical modelling approach to intra-party competition under preferential voting rules have already been attempted, most notably in a paper by Shugart, Bergman and Watt (2013), henceforth SBW, in the context of a comparison between open-list and single non-transferable vote systems (a slightly modified version is in Shugart and Taagepera, 2017, 215-235). The basic building block of the SBW model is the formula for the fractional share of preference votes for the first-ranked[4] candidate of a seat-winning list. An appealing feature of this model is its parsimony: in its simplest form, the SBW models employs only one input variable, c, the number of candidates in a list. However, the deriving predictions fall well short of the accuracy of those available for the inter-party dimension, and to improve accuracy SBW often have to rely on constants derived empirically that correct the equations for biases that cannot be accounted for theoretically. Moreover, these models' scope has so far been limited to a relatively narrow set of 'simple' preferential voting systems, which exclude flexible list types or systems that allow multiple preference votes. The wager of this paper is that we can do better at relatively little cost in terms of simplicity.

It is however worth revisiting the derivation of v_1 in the SBW model, as an illustration of how logical

[4]I used the term 'rank' to refer to the ordering of candidates according to their preference votes, and 'position' to denote their order on the ballot paper.

model-building proceeds. The first step consists in identifying the conceptual boundaries of the quantity of interest. At one extreme, v_1 may not exceed 1: at best, the first-placed candidate can get *all* the preference votes cast for a given list. As for the lower bound, v_1 may not be lower than $\frac{1}{c}$: this is the case where all candidates in the list get the same share of votes, so that the list is maximally competitive. A candidate getting less than $\frac{1}{c}$ cannot logically come first in the list. The expectation for v_1 lies between these two bounds and can be approximated as an average of the two. In line with the rest of the literature on logical models of electoral system quantities, the geometric mean is preferred to the arithmetic mean,[5] so that

$$v_1 = \left(1 \times \frac{1}{c}\right)^{\frac{1}{2}} = c^{-\frac{1}{2}} \tag{5.1}$$

That is, the predicted fractional share for the first candidate in a list with c candidates is the inverse of the square root of c.

5.2.2 The Revised Model

This SBW model is indeed the 'best guess' for v_1, when this quantity is hypothesised to depend uniquely on c. Moreover, it is derived from 'hard' conceptual boundaries, which are *impossible* for v_1 to cross. Let us now relax these conditions by considering the role of two additional input variables: the number of preference votes and the expected number of seats at stake. Essentially, this new version of the model attempts to incorporate two intuitions in the derivation of v_1:

- The competitiveness of a list should increase in the number of preference votes that each voter can cast. Thus v_1 should decrease as voters are allowed to express more preferences (provided that they are non-cumulative).
- The competitiveness of a list should increase in the expectation of the number of seats attributed to each list. Thus v_1 should decrease as lists are expected to elect more candidates.

To model these assumptions, we proceed by replacing the hard conceptual bounds in equation 5.1 with some

[5]Logical models effectively approximate the *median* value of a quantity, returning predictions that are in expectation equally likely to be below or above its actual value. Geometric means between conceptually extreme cases express the median of distributions better than arithmetic means when the variable of interest can only take positive values, and therefore a normal distribution cannot be assumed. In particular, the geometric mean is preferable when the assumed distribution of a variable spans different orders of magnitude, as it abides by the principle of equal distortion: when we have no priors over two extreme possible values x_{max} and x_{min}, a 'best guess' that is equally likely to be above or below the true value is that both are off by the same multiplicative factor k, so that $x^* = (x_{max} \times x_{min})^{\frac{1}{2}}$. In the specific case of the coarse model for v_1, the arithmetic mean of conceptual boundaries would predict a first-ranked candidate share of 0.5 in the limit for increasing values of c which is an implausible degree of concentration for an extremely high number of candidates. For more details on the use of geometric means in logical models see Taagepera (2008, pp. 120–124).

'soft' upper and lower bounds, beyond which the quantity is *unlikely* to lie due to the expected effect of the additional input variables. Because these new boundaries are themselves derived from the geometric mean of conceptual bounds, the revised model will therefore be expressed in terms of 'predictions from predictions'.

The Number of Preference Votes (p)

Many preferential-list systems allow voters to express one or more non-transferable and non-cumulative[6] preferences for different candidates: this is the case in a few open-list (Cyprus, Greece, Italy, Kosovo, Peru, Sri Lanka) and many flexible-list systems (Belgium, Bosnia, Czechia, Slovakia). This factor is important because *the number of preferential votes at voters' disposal can be expected to compress the upper bound on the share of preference for the first party.* For instance, consider a case where voters *must* cast two preferences. Even if a candidate gets a vote from *all* electors, her share of preference votes will not cross $\frac{1}{2}$, as voters' second preference is spread across the other candidates in the list. To my knowledge, no preferential voting system employed for a parliamentary election requires voters to cast more than one preference vote; however, many *allow* such an option. In this case, the 'hard' upper bound for v_1 in a system where voters may cast from 1 to p preference votes is still 1: the scenario where all voters cast one preference and they all go to one candidate. But we may identify a more realistic – or at least, more useful – 'soft' upper bound by positing that, in such a system, voters will cast a number of preferential votes comprised between 1 and p. If all voters cast one vote, then the maximum fractional share the first-ranked candidate can obtain is still 1; if all voters cast p votes, then it is $\frac{1}{p}$. Taking the geometric mean of these boundaries, therefore our 'soft' upper bound for v_1 is $\frac{1}{p^{\frac{1}{2}}}$, or $p^{-\frac{1}{2}}$, where p is the maximum number of preference votes at voters' disposal. Of course, the logical standing of this boundary condition is only as good as our assumption of the number of preference votes cast by each voter: if they cast more, it should be lower; if they cast fewer, it should be higher. In this sense, it is a 'prediction from prediction', rather than a conclusion from pure deductive reasoning.

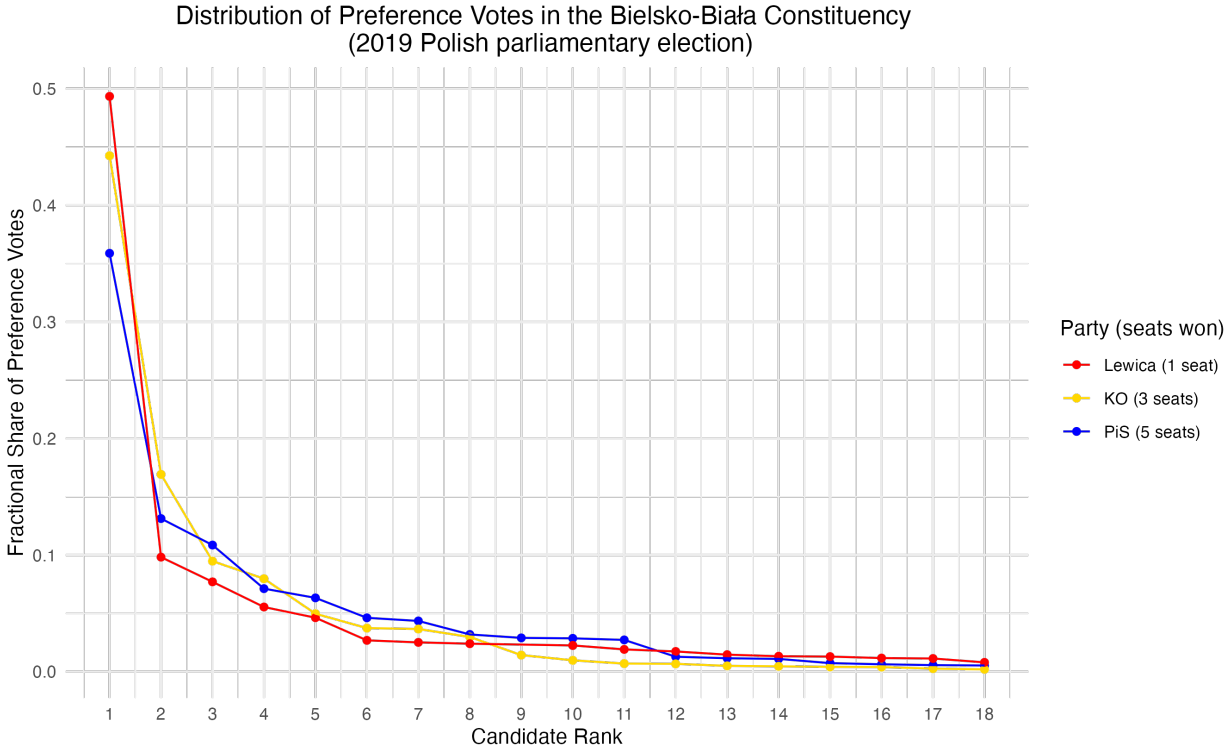
Expectations of Inter-party Performance

The second modification of the SBW model attempts to model the assumption that *intra-party competition in a list is endogenous to actors' expectation of the list's inter-party performance.* In other words, assuming that vote-seeking is costly, the more seats are expected to be assigned to a list, the more candidates will be

[6]There are rare cases that allow voters to express multiple preferences for the same candidate, such as in Switzerland, Luxembourg, and Hamburg's and Bremen's State Parliament electoral systems. Because non-cumulability of the vote is a crucial assumption at this stage, such preferential-list systems fall outside of the scope conditions of the model.

serious about seeking preference votes.[7] In a way, this assumption simply maps Duverger's intuition onto intra-party competition: just as the number of electorally competitive parties decreases with the district magnitude M, the number of electorally competitive candidates decreases in the expectation of the number of seats at stake s.

Figure 5.1: Preference votes in the Bielsko-Biała constituency (2019 Polish parliamentary election).



This point is perhaps best illustrated by peeking (temporarily) behind the veil of ignorance-based theorising and into real-world results. Figure 5.1 shows the distribution of preference votes for the three seat-winning parties competing in the Bielsko-Biała (Nr. 27) constituency in the 2019 Polish parliamentary election.[8] The first thing to note from a glance at the plot is that most candidates for all parties get close to zero percent of preference votes (not an uncommon pattern: see, for instance, Cheibub and Sin, 2020). Most names appearing on the ballot are effectively 'list fodder' or 'top-up candidates' (Arter, 2013), with little personal support or hope for election. This observation suggests that c, the key variable in the SBW model, can be a very volatile and imprecise predictor of intra-party competition: if the parties had only fielded 9 candidates

[7] Another way of thinking about this is from voters' perspective: the more seats are expected to be assigned to a list, the less voters will be concerned about wasting their preference on non-viable candidates.

[8] It can be shown empirically that v_1 is significantly lower for more successful seat-winning parties, discounting all district-level variation through fixed effects. See Appendix section G.

instead of 18, the SBW model's prediction would change dramatically (the upper bound would double and v_1 would increase by a factor of $\sqrt{2}$), whereas in fact the last-placed 9 candidates account for only 10% of each party's total share of preference votes.

Secondly, the plot shows that the list that came on top in the district – the right-wing *Prawo i Sprawiedliwość* (PiS), which won five seats – has a lower first-candidate preference share of the vote and a flatter distribution of the vote than the other two, the liberal *Koalicja Obywatelska* (KO) and the left-wing *Lewica Demokraticzna* (Lewica). This observation is consistent with the idea that the number of electorally competitive candidates behaves with respects to the expectations of a list's performance in a similar way to how the number of electorally competitive parties behaves with respects to the district magnitude: the fewer spots available, the fewer serious contenders. (And, in turn: the fewer serious contenders, the more concentrated preferential votes will be towards the top-end of the candidate ranking.)

This intuition is perhaps the main innovative aspect of this paper – though Crisp, Jensen and Shomer (2007) made a similar observation – but its formalisation requires something of a leap of faith. How to model the effect of seat expectations on intra-party competition? My proposal is to introduce the concept of 'number of pertinent vote-earning candidates' c^* as a substitute for c, in order to formalise the idea that using simply the list length c will fail to differentiate how lists with higher or lower seat-winning potential will have different levels of internal competitiveness. Conceptually, c^* is an analogue of N_{v_0}, the number of pertinent vote-earning parties in a district (Shugart and Taagepera, 2017, p. 128): a quantity that denotes "how many [parties] are sufficiently important to contribute to our prediction." In the same way, c^* is a 'phantom quantity': it does not refer to an observable, but it helps to derive one, because it incorporates the idea that fewer candidates will seriously compete for votes when fewer seats are expected to be a stake. In this sense, a list will be *de facto* shorter when candidates expect there to be few seats at stake, and *de facto* longer when candidates are competing for more spots. Moreover, c^* is distinct from, say, the effective number of candidates, because we are not mainly interested in the size of their preference vote share: holding c constant, a candidate getting 1% of preference votes may be hopeless in a list competing for one seat, but has a realistic shot at a seat in a list competing for 15 spots in parliament.

Because expectations are crucial to these theoretical steps, the conceptual boundaries of the number pertinent vote-earning candidates c^* are identified in terms of actors' ability to infer list performance. On the one hand, candidates may be entirely in the dark about their electoral potential and the realistic number of seats allocated to their list. In this case, they all vie for votes, so that $c^* = c$, as in the SBW model. At the

other extreme, candidates may have perfect foresight about election results and their vote-winning potential, so that only the candidates who expect to end up in seat-eligible spots in the final rank order campaign and vie for votes.[9] In this case, where actors are perfectly efficient in choosing their level of campaign effort and have perfect priors over the number of seats the party will win, $c^* = s$, where s notates the number of seats won by the party in the district. Hence the expected number of pertinent vote-earning candidates in a list of size c gaining s seats can be computed as the geometric mean of s and c: $c^* = (sc)^{\frac{1}{2}}$.[10] It follows that the revised lower bound for v_1 is the inverse of c^*: considering only pertinent vote-earning candidates, at its lowest v_1 takes the value of $\frac{1}{c^*} = (sc)^{-\frac{1}{2}}$, when all the 'serious' candidates get the same share of the vote.

Bringing It All Together

Substituting the soft upper bound $p^{-\frac{1}{2}}$ and the soft lower bound $(sc)^{-\frac{1}{2}}$ into equation 5.1, we may therefore derive the revised prediction for the value of v_1:

$$v_1 = \left(\frac{1}{(sc)^{\frac{1}{2}}} \times \frac{1}{p^{\frac{1}{2}}}\right)^{\frac{1}{2}} = (scp)^{-\frac{1}{4}} \tag{5.2}$$

To recapitulate, equation 5.2 expresses that the fractional share of preference votes obtained by the first-ranked candidate v_1 is comprised between these two bounds:

— The candidate's share under the least-competitive scenario, where the the first candidate gets one preference from *all* voters, and voters on average cast $p^{\frac{1}{2}}$ preferences each, with p being the maximum number allowed. For instance for $p = 2$, 100 voters can be expected to cast a total of approximately 141 votes, so that the upper bound for v_1 is $100/141 \approx 0.71$ or equivalently $\frac{1}{2^{\frac{1}{2}}} \approx 0.71$.

— The candidate's share under the most-competitive scenario, where all pertinent vote-earning candi-

[9] It is not unrealistic that candidates have decent priors over their list's inter-party competitiveness; as Shugart and Taagepera (1989, p. 215) put it, "in a multi-seat district with a fairly stable voting pattern, the number of seats one particular party can obtain is known ahead of time within plus or minus one seat." Furthermore, candidates may have decent *ex ante* information about *their own* vote-winning potential as well, as their ballot position is strongly (and, in part, causally) associated with their chances (Lutz, 2010; Blom-Hansen et al., 2016; Devroe and Wauters, 2020; Van Erkel and Thijssen, 2016).

[10] To be exact, the model uses s, the observed number of seats won, as a proxy for $E(s)$, the unobservable number of seats actors in a list *expect* to win. This requires us to restrict the analysis to 'seat-winning' parties, discounting unsuccessful lists, where s is zero. c^* may be adjusted to be $[(s+1) \times c]^{1/2}$ – i.e. assuming that lists are at least marginally over-optimistic about their potential, otherwise they would not compete – to predict preference shares of non-seat-winning parties as well. A similar adjustment is proposed by Selb and Lutz (2015, p. 332): "we rely on the assumption that candidates' expectations are correct on average, and use the actual number of seats won by the candidates' list in the current election as a proxy for the expected number of seats. Moreover, to avoid substantial losses of observations due to zero divisions for all the lists that did not gain any seats, we will add one seat to the denominator [...] which implements the reasoning that any list which stands for election does so because its members expect at least one seat." However, to avoid the complication of introducing a further constant, the main model retains the simpler formula, and keeps the focus on seat-winning parties for reason of substantive interest and in line with previous theoretical work. Appendix section H.1 reproduces the analysis in section 5.6.1 with predictions based on a lower bound for c^* set to be $[c(s+1)]^{1/2}$ (on the same sample of seat-winning lists used the main analysis).

dates, a quantity endogenous to the 'length' of a list and its expected performance, get the same number of votes each. For instance, in a list with 25 candidates expected to gain 6 seats, the number of pertinent vote-earning candidates is $(25 \times 6)^{\frac{1}{2}} \approx 12.25$, and minimum value of v_1 is $\frac{1}{12.25} \approx 0.08$.

Thus, in this example, our best guess for v_1 in a list with 25 candidates winning 6 seats under PLPR rules allowing a maximum of 2 preference votes is $(2^{-\frac{1}{2}} \times (25 \times 6)^{-\frac{1}{2}})^{\frac{1}{2}} \approx (0.71 \times 0.08)^{\frac{1}{2}} \approx 0.24$. Note that, because $s \geq 1$ (the scope conditions are limited to seat-winning parties) and $c \geq p$ (voters cannot cast more votes than there are candidates), it follows that $(sc)^{-\frac{1}{2}} \leq p^{-\frac{1}{2}}$: the 'soft' upper bound may not logically be lower than the 'soft' lower bound.

Unmodelled Variables

Although a three-variable model is perhaps already complex enough to test Taagepera and colleagues' injunction to make logical models "as simple as possible, but no simpler" (Taagepera, Selb and Grofman, 2014, pp. 396-397, attributed to Albert Einstein), it obviously does not exhaust all possible sources of variation in intra-party competitiveness. Ultimately, whether and how well one can really predict average values of v_1 from variation in s, c and p is an empirical question. But it is worth briefly mentioning what has been left out of the picture.

First, the model makes no reference to the type of PLPR it applies to, the relevant distinction here being between open and flexible lists. While in the former all candidates are elected according to the order of preferences received, the latter require candidates to reach some preference threshold to be elected via preferences, while the rest of the seats are filled according to the party list position. In short, we simply have no directional prior as to how list type should matter, let alone a way to quantify such an effect. On the one hand, a flexible list system may decrease first-candidate vote share, if candidates listed in higher positions put less effort in attracting preferential votes, knowing they are likely to be elected regardless. On the other hand, a flexible list system might make the distribution of votes steeper, dissuading lower-positioned candidates from vying for votes, as those may be insufficient to clear the threshold.[11] Secondly, the model does not distinguish between contexts where preferential vote is mandatory and those where it is optional. Once again, there is no clear directional prior as to the effect of this variable, and effectively we should take a leap of faith and assume that the model predictions will apply *on average* across a sample of diverse

[11] It is however worth noting that, as flexible-list systems normally allow multiple preference votes, having included p as an input variable makes it more realistic that the scope conditions can be extended beyond simple open-list PR. The empirical analysis in section 5.6.1 includes separate tests of the models for open- and flexible-list elections.

institutional contexts. Factors falling outside the realm of electoral institutions – larger parties may be more factionalist, smaller parties may have fewer qualified candidates, outcomes are more easily predictable in stable party systems – are also beyond the reach and remit of a logical model.

5.3 Extending the Intra-party Model

This section develops the fundamental building block of the model, the equation $v_1 = (scp)^{-\frac{1}{4}}$, in two directions:

— In subsection 5.3.1, it is shown that two additional quantities of interest describing the degree of intra-party competitiveness of a list can be derived from v_1: the effective number of candidates (N_c) and the vote share for the last eligible candidate (v_s). First, I show how these quantities can be computed via an algorithmic iteration of the same procedure used to derive v_1 that derives predictions for preference shares for *all* candidates in a list. Then, I propose some approximations for these quantities that retain the simple X^k structure of the formula for v_1.

— In subsection 5.3.2, I recast the formula for v_1 – and the related equations for N_c and v_s – in terms of variables that are purely institutional, and can be gathered independently of outcomes realised on election day. This yields a 'results-blind', institutions-only model that makes predictions for average expected values of these quantities *at district level*, as opposed to the list-level prediction of the model in terms of s, c and p.

5.3.1 Deriving N_c and v_s

Algorithmic Approach

Just as we did with v_1, we may use the same process of individuating conceptual boundaries and taking their geometric mean to derive predictions for the values of $v_2, v_3, v_4 ... v_c$. These values would allow to derive N_c and v_s from the distribution of expected preference vote shares for *all* candidates. As we will see, this is algebraically messy. However, it is worth outlining how we may obtain these quantities with an iterative algorithm, as these predictions can serve as 'sanity checks' for the more synthetic approximations of N_c and v_s described in the rest of this section.

Let us start from v_2: the fractional share of preferential votes for the second-ranked candidate. Once v_1 is derived, we may use an analogous logic to derive the vote share of the second elected candidate, v_2. We

must, however, distinguish two cases.

- **Case 1** $v_1 \leq 0.5$

If the first-ranked candidate gets less than half of the preference votes, the upper bound for v_2 is v_1: the second candidate may not get more votes than the first candidate. The lower bound corresponds to the case where all the remaining pertinent vote-earning candidates get an equal share of the remaining vote share once v_1 is realised. The lower bound for v_2 is therefore $\frac{1-v_1}{(s\times(c-1))^{\frac{1}{2}}}$. The geometric mean of lower and upper bounds is the expected share of votes for the second candidate elected:

$$v_2 = \left(v_1 \times \frac{1-v_1}{(s\times(c-1))^{\frac{1}{2}}}\right)^{\frac{1}{2}} \tag{5.3}$$

- **Case 2** $v_1 > 0.5$

We must however consider the case in which the first candidate gets *more* than half of the preference votes. Such circumstance makes it illogical to posit v_1 as the upper bound for the second candidate: the second candidate's share never be as much, as the sum of first and second candidate shares would exceed 1. In this case, we must substitute in equation 5.3 the *actual* upper bound, which will be $1 - v_1$:

$$v_2 = \left((1-v_1) \times \frac{1-v_1}{(s\times(c-1))^{\frac{1}{2}}}\right)^{\frac{1}{2}} = \frac{1-v_1}{(s\times(c-1))^{\frac{1}{4}}} \tag{5.4}$$

Equations 5.3 and 5.4 can be generalised to the n^{th} candidate, whose share v_n is the geometric mean between the following two conceptual bounds:

— the upper bound is whichever is smaller of v_{n-1} (the share of the $(n-1)^{th}$ candidate) and $1-\sum_{i=1}^{n-1} v_i$ (the remaining share of the vote after v_{n-1} is realised). That is, the share of the vote for the n^{th} candidate is subject to the conditions that it cannot be greater than the share of the vote of the candidate placed above her and it cannot tip the total of preference shares above 1.

— the lower bound is $\frac{1-\sum_{i=1}^{n-1} v_i}{(s(c+1-n))^{1/2}}$, i.e. 1 minus the sum of $v_1, v_2, v_3 ... v_{n-1}$ over the number of pertinent vote-earning candidates $(s(c+1-n))^{\frac{1}{2}}$, where c is reduced by one each time one candidate share is realised. That is, the share of the vote for the n^{th} candidate is smallest when she gets the same preference votes as all the pertinent vote-earning candidates placed below her.[12]

[12]To compute values for candidates beyond c^* and allow the sum of all the fractional shares to converge to 1 in the limit, we need to hold s constant rather than using $(s+1-n)$ as we did with c. In this way, the number of candidates that are relevant to compute for second, third, fourth etc. place is reduced at each iteration – because the list gets shorter once each fractional share is realised – but never falls below zero.

The generalised algorithm to compute v_n $(2 \leq n \leq c)$ is therefore as follows:

$$v_n = \left(min\{v_{n-1}, 1 - \sum_{i=1}^{n-1} v_i\} \times \frac{1 - \sum_{i=1}^{n-1} v_i}{(s(c+1-n))^{1/2}} \right)^{\frac{1}{2}} \tag{5.5}$$

An Approximation of N_c

Having derived $v_1, v_2, v_3 ... v_c$ with the algorithm in equation 5.5, we may therefore compute the effective number of candidates N_c from its definition as $N_c \equiv \frac{1}{\sum_1^c ({v_i}^2)}$. This obviously does not simplify to a neat, generalisable expression in the form of X^k, or at least not one with a tractable value of the base. We may therefore attempt to use a shortcut to derive N_c from approximated conceptual boundaries expressed in terms of v_1. To distinguish this 'approximated' prediction for N_c from the value of the effective number of candidates obtained from iterating the algorithm in equation 5.5, I temporarily note the approximation as $\hat{N}_c$, with a hat.

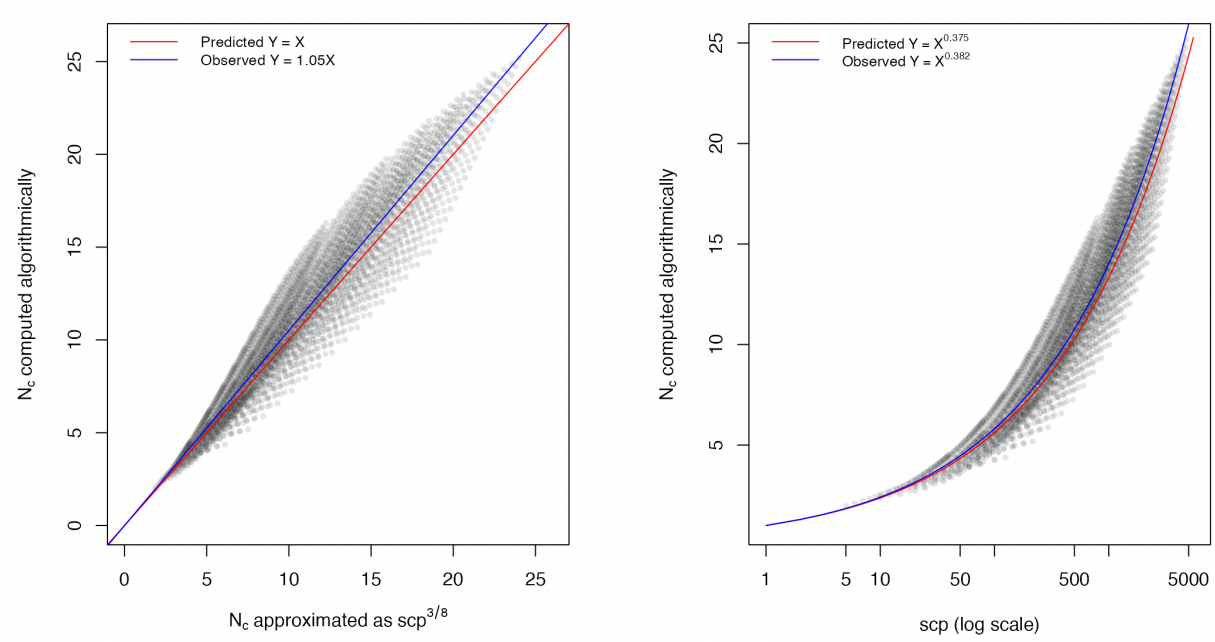
The value of the effective number of candidates N_c is at its greatest when all candidates share the remaining share of the vote after v_1 is realised equally. For a sufficiently large value of c, these small shares, corresponding to $\frac{1-v_1}{c-1}$ each, will become negligible when squared. Hence, the upper bound can be approximated as $\frac{1}{{v_1}^2+0^2+0^2...} = \frac{1}{{v_1}^2}$. The lower bound for v_1 corresponds to the case where the distribution of the vote is as 'compact' as possible: because the maximum a candidate can get is v_1, there will be $\frac{1}{v_1}$ candidates getting v_1 share of the vote each. It follows that the lower bound will be approximately $\frac{1}{\frac{1}{v_1} \times v_1^2} = \frac{1}{v_1}$.[13] Taking the geometric means of these conceptual bounds and substituting our v_1 equivalence from equation 5.2, we obtain the following approximation:

$$\hat{N}_c = \left(\frac{1}{v_1} \times \frac{1}{{v_1}^2} \right)^{\frac{1}{2}} = {v_1}^{-\frac{3}{2}} = (scp)^{\frac{3}{8}} \tag{5.6}$$

It can be shown graphically that this estimation approximates quite well the value of the effective number of candidates N_c computed via algorithmic iteration, at least for realistic values of s, c and p. I simulated data with all combinations of c comprised between 5 and 40, s comprised between 1 and 20, and p comprised between 1 and 6. After removing the 'illogical' cases where $s > c$ or $p > c$, I computed N_c with the algorithmic method for all combinations of values and $\hat{N}_c$ with the approximation in equation 5.6. The graph on the left in figure 5.2 shows how the two estimates compare. Fitting a fixed-intercept linear regression $(Y = \beta X)$ returns a linear relationship between the two corresponding to $N_c = 1.05(\hat{N}_c)$, which is reasonably close to

[13] A similar procedure to approximate the effective number of parties is in Taagepera and Shugart (1993).

Figure 5.2: Comparison of N_c estimates.



the expected identity $N_c = \hat{N}_c$. The graph on the right in figure 5.2 shows the relationship between the 'raw' measure of the *scp* product and the algorithmic estimation of the effective number of candidates N_c. Modelling the relationship as a fixed exponent regression ($Y = X^{\beta}$) returns the function $N_c = scp^{0.382}$, which again is closely in line with our approximation $\hat{N}_c = scp^{\frac{3}{8}}$. To be sure, this does not mean the approximation for the effective number of candidates is 'true': it simply means that it is consistent with the generative process assumed to be behind the estimation of v_1 and extended via algorithmic iteration to $v_n \mid n \geq 2$.

An Approximation of v_s

A free-standing, empirically useful formula for v_s has remained elusive to logical modellers, to the extent that the two existing attempts at deriving this quantity (Shugart, Bergman and Watt, 2013; Shugart and Taagepera, 2017, pp. 226-235) ultimately recur to empirically derived constants to adjust the models. In this section, I propose a new approach to derive a 'shortcut' approximation for v_s. The method proposed is somewhat more convoluted than the approximation of N_c and involves something of a mathematical sleight of hand in positing the mean and median of a quantity to be approximately equal in expectation. Nonetheless, as it will be shown, the deriving formula, when expressed in purely institutional terms without empirical

input, performs respectably well on both simulated and real-world data.

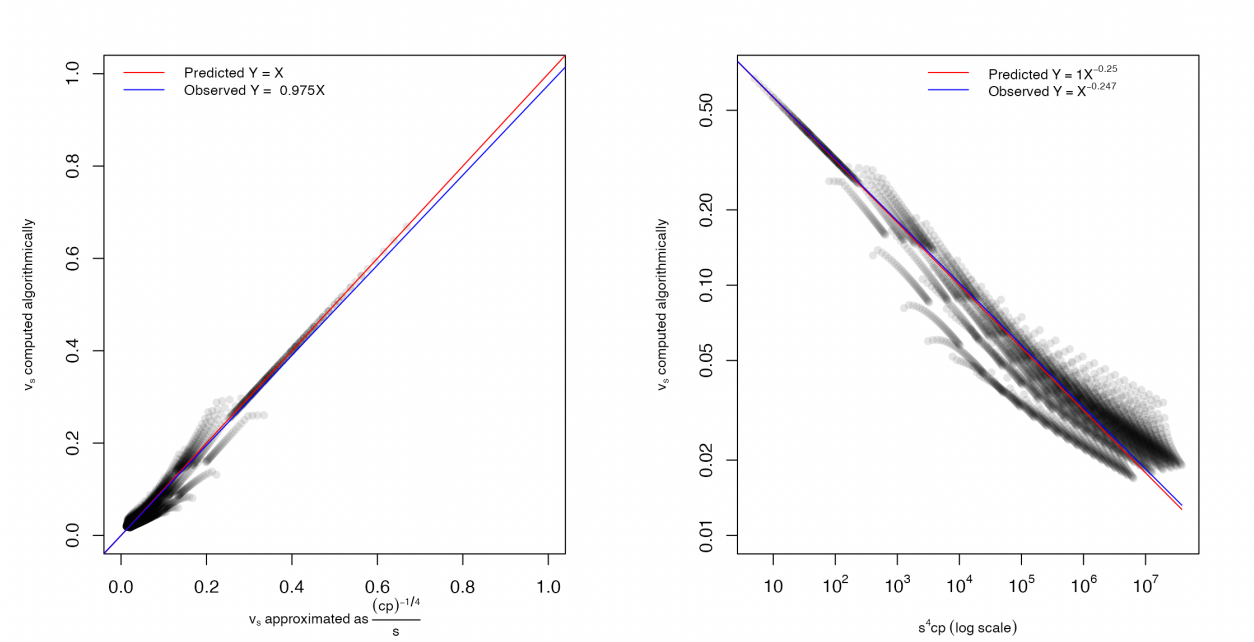
Let us start by considering the *expected share for a seat-eligible candidate*, $v_{i\leq s}$. We can estimate such a quantity as a function of v_1 (in expectation) in two ways: either as the expected median of preference shares of seat-eligible candidates (notated with a tilde $\tilde{v}_{i\leq s}$) or as the expected mean of preference shares of seat-eligible candidates (notated with a bar $\bar{v}_{i\leq s}$). First, relying on Taagepera's argument that the geometric mean of conceptual bounds approximates a median, we can derive the expected median preference vote for a seat-eligible candidate $\tilde{v}_{i\leq s}$ as a quantity comprised between v_1 and v_s, and therefore $\tilde{v}_{i\leq s} = (v_1 \times v_s)^{\frac{1}{2}}$. Let us now repeat the conceptual boundaries logic for the mean preference share for a seat-eligible candidate $\bar{v}_{i\leq s}$, starting from a computation of upper and lower bounds for the mean as functions of the realised value of v_1. The mean preference share is at its highest when $s = 1$, the list only wins one seat and the mean is v_1: hence, $(1 \times c \times p)^{-\frac{1}{4}}$ is the upper bound. The lower bound is realised when all candidates after v_1 up to v_s get as few preference shares as possible: for values of c sufficiently larger than s, the value of each individual one of these $\frac{1-v_1}{c-1}$ shares will tend to zero, so that the expected mean preference vote will be approximately $\frac{v_1}{s}$. We can thus proceed to derive the mean preference vote of seat-eligible candidates in expectation as $\bar{v}_{i\leq s} = ((cp)^{-\frac{1}{4}} \times \frac{v_1}{s})^{\frac{1}{2}}$.

An approximation for v_s, notated as $\hat{v}_s$, can be derived algebraically by positing that the median ($\tilde{v}_{i\leq s}$) and mean ($\bar{v}_{i\leq s}$) share of a seat-eligible candidate's preference votes are approximately equal:

$$\begin{aligned} \tilde{v}_{i\leq s} &\approx \bar{v}_{i\leq s} \\ (v_1 \times v_s)^{\frac{1}{2}} &\approx \left((cp)^{-\frac{1}{4}} \times \frac{v_1}{s}\right)^{\frac{1}{2}} \\ v_s &\approx \frac{(cp)^{-\frac{1}{4}}}{s} \\ \hat{v}_s &= (s^4 cp)^{-\frac{1}{4}} \end{aligned} \tag{5.7}$$

Let us now repeat the same 'sanity checks' for $\hat{v}_s$ as we did for $\hat{N}_c$ (graphically, in figure 5.3). Regressing the algorithmically derived value of the share for the last eligible candidate v_s on the approximation $\hat{v}_s$ obtained from equation 5.7, we obtain $v_s = 0.975\hat{v}_s$, which is close to an identity. The exponential function linking v_s to s^4cp is $v_s = (s^4cp)^{-0.247}$, which is close to the the approximation's prediction $\hat{v}_s = (s^4cp)^{-\frac{1}{4}}$. Even so, the theoretical derivation of this approximation is admittedly not as rigorous or satisfying as we would wish, as it relies on the assumption that two quantities that we *know* to be distinct are approximately

Figure 5.3: Comparison of v_s estimates.



equal. In particular, the comparability of the two measures of central tendency may be limited when parties gain many seats, and thus the distribution of preferences for seat-eligible candidates becomes more skewed. There are hints of this bias in figure 5.3, which suggests that for very high values of the base product (low values of v_s), the approximation somewhat over-estimates the quantity relative to the algorithm's prediction.

5.3.2 Predicting v_1, N_c and v_s from Institutional Variables

The equations $v_1 = (scp)^{-\frac{1}{4}}$, $N_c = (scp)^{\frac{3}{8}}$ and $v_s = (s^4cp)^{-\frac{1}{4}}$ are 'list-level' models insofar as they predict different values for observations in the same district, varying from list to list as a function of their different values of s and c (conversely, p is constant in a district).[14] However, as s and c are only realised at election time, these models do not fully qualify as *pre*-dictive: they can only yield predictions after candidate lists are presented by parties and inter-party seat allocation is realised. To be of use to institutional engineers, we need to recast these models in terms of average expected indicators of intra-party competitiveness that vary as a function of variables that pre-exist the selection and election of candidates. In particular, as it will be argued shortly, both s and c can be expressed as some function of the district magnitude. In doing

[14] I henceforth refer to $\hat{N}_c$ and $\hat{v}_s$, the approximations of the quantities, simply as N_c and v_s, as the algorithm-based prediction for these quantities are not relevant to the rest of the analysis.

so, the models will shift from the list level to the district level: if we do not know the exact values of c and s for each list, we may only express these quantities in terms of their expected value for *any* seat-winning list in a district. In practice, this is the same as deriving the value of a quantity for the expected median seat-winning party, as this prediction would thus be equally likely to fall below or above the real quantity for any party.

First, let us consider how c is related to electoral institutions. How many candidates can a seat-winning list be expected to field in a district where M seats are at stake? Here we need to make a further assumption based on observed empirical regularities: *lists nominate as many candidates as they are legally allowed to*. This is because, unlike under STV and SNTV, preferential-list systems *pool* individual candidates' votes into list totals, which in turn determine the inter-party allocation of seats. Thus, the returns to additional candidate nominations are always positive for parties, as there is no risk of 'wasting' votes through over-nomination error (Shugart, Bergman and Watt, 2013; André, Depauw and Deschouwer, 2014). Legal limits to the number of candidates vary from electoral system to electoral system. In countries like Italy, Cyprus and Belgium, the number of candidates in a party-list is capped to the number of seats at stake in a district, so that $c = M$. Elsewhere, they might nominate up to $M + 2$ (Estonia) or $2M$ (Poland). Therefore, in general, $c = rM$, where r is the ratio between the maximum number of candidates allowed and the district magnitude, varying at district level. Having assumed that seat-winning lists always over-nominate, our 'over-nomination ratio' parameter r is thus a purely institutional measure, albeit one that is often hidden in obscure electoral regulations.[15]

Secondly, let us consider how s may be expressed as a function of institutional quantities. In this case, we once again employ the prediction-from-prediction approach to derive the average (in this case, median) expected number of seat for any seat-winning party in terms of M. The work of Taagepera and Shugart (1993) shows that the number of seat-winning parties N'_0 in a district is $M^{\frac{1}{2}}$: the geometric mean between the minimum of 1 and the maximum of M. From here, they derive the fractional share of seats σ'_1 won by the largest party *in a district* as the geometric mean between the maximum of 1 and the minimum of $\frac{1}{N'_0}$, returning $\sigma'_1 = M^{-\frac{1}{4}}$. The expected number of seats won by the largest party is therefore $M \times \sigma'_1 = M^{\frac{3}{4}}$.[16]

[15] This assumption holds up well empirically for most seat-winning parties (see section 5.4). Very small lists may be short of personnel and thus be forced to under-nominate; but as our focus is on seat-winning parties, these lists are unlikely to be relevant. A thornier problem is represented by countries with particularly complex over-nomination rules. For instance, in Brazil, each *individual party* in a *coalition* list may nominate up to M candidates. This means not only that r varies as a function of 'political' factors, rather than simply institutional ones, but also that some parties may prefer to under-nominate to concentrate their preference votes within the coalition. Such a case is therefore not tractable in an institutions-only model.

[16] In a more recent version of their predictive model of district-level party shares (Shugart and Taagepera, 2017, 153–180), they introduce a constant k to adjust the exponent of σ'_1 (and the deriving quantities) for the 'embeddedness' of districts in the broader political system. The deriving model predictions for intra-party quantities and their performance are very similar

It follows that the expected number of seats for *any* seat-winning party is the geometric mean between the minimum of 1 and the maximum of $M^{\frac{3}{4}}$, i.e. $M^{\frac{3}{8}}$.

Substituting rM for c and $M^{\frac{3}{8}}$ for s, we obtain district-level predictions for our quantities of interest in terms of M, r, and p. These should predict the expected variation in the dependent variables across *any* seat-winning list.

$$v_1 = (scp)^{-\frac{1}{4}} = (prM^{\frac{11}{8}})^{-\frac{1}{4}} \tag{5.8}$$

$$N_c = (scp)^{\frac{3}{8}} = (prM^{\frac{11}{8}})^{\frac{3}{8}} \tag{5.9}$$

$$v_s = (s^4cp)^{-\frac{1}{4}} = (prM^{\frac{5}{2}})^{-\frac{1}{4}} \tag{5.10}$$

Of course, we need not pass by the list-level model to derive these district-level predictions (although it is a useful illustrative step). For instance, equation 5.8 can be obtained as the geometric mean of the upper bound $p^{-\frac{1}{2}}$ and the lower bound $(M^{\frac{3}{8}} \times rM)^{-\frac{1}{2}}$. The lower bound is in turn $\frac{1}{c^*}$, i.e. the inverse of the geometric mean between rM, the maximum number of candidates a list can field, and $M^{\frac{3}{8}}$, the expected number of seat-winners for the median seat-winning list, denoting the minimum number of candidates that have an incentive to compete for votes in such a list.

5.4 Data

To test the models, I collected preference shares for candidates at the district-list-election level for 31 Lower Chamber elections in nine preferential-list PR systems, including both open- and flexible-list systems.[17] Preference shares for each candidate in a list are defined as a candidate's share of all preferences cast for the party in the district they run in. From these, I coded the actual values of v_1, N_c and v_s for each list-district-election observation: these serve as the dependent variables in the empirical test of the models.

All these elections were conducted under PLPR rules, and there was no major institutional change over the period of time considered. Table 5.1 summarises some key inter-party and intra-party characteristics

using this more complex value: see Appendix section H.2.

[17] I excluded results from two single-member districts: Aosta, where $c = 1$ and thus is effectively not a preferential-list system; and Åland, where the data available does not allow to isolate which candidates belong to which lists (all are listed as 'Other' as the Swedish-speaking region has its own party system).

Table 5.1

General Information

Country	Years	Assembly Name	Elections	Districts	Size	list type	Pref. Vote
Belgium	2003-19	Chambre des Représentants	5	11	150	flexible	optional
Cyprus	2011-21	Voulí ton Antiprosópon	3	6	56	open[a]	optional
Czechia	2013-21	Poslanecká Sněmovna	3	14	200	flexible	optional
Estonia	2011-19	Riigikogu	3	12	101	flexible	mandatory
Finland	2011-19	Eduskunta	3	12	200	open	mandatory
Italy	1976-92	Camera dei Deputati	5	31	630	open	optional
Peru	2014-21	Congreso de la República	3	26[b]	130	open	optional
Poland	2011-19	Sejm	3	41	460	open	mandatory
Slovakia	2012-20	Národná Rada	3	1	150	flexible	optional

Seat Distribution Rules

	Inter-party dimension		*Intra-party dimension*		
Country	PR Formula	Party Thresh.	Max. No. Cands.	Preference Thresh.	Maximum Votes
Belgium	D'Hondt	—	M	quota-based[c]	M
Cyprus	Hare	3.6%	M	—	$M/4$ (rounded up)
Czechia	D'Hondt	5%	varies by district	5% of party votes	4
Estonia	D'Hondt[d]	5%	M+2	quota-based[c]	1
Finland	D'Hondt	—	M or 14 (for $M \leq 14$)	—	1
Italy	Imperiali[d]	—	M	—	3 or 4 (for $M \geq 15$)
Peru	D'Hondt	5%	M or 3 (for $M \leq 3$)[e]	—	2
Poland	Sainte-Laguë	5%[f]	$2 \times M$	—	1
Slovakia	Hag.-Bischoff	5%	M	3% of party votes	4

Notes: (a) In Cyprus, the vast majority of MPs are elected via open-list PR. However, party leaders are elected automatically from the list they run in. But this applies only to 5-6 candidates per election-year, so while technically flexible, the system might be considered virtually open (Passarelli, 2020, pp. 92-93). (b) In the 2021 election, there were 27 districts, as the district reserved for voters resident abroad was separated from the Lima district. (c) In Belgium, the threshold is equal to the party's Hagenbach-Bischoff quota; in Estonia, the threshold is 10% of the Hare quota. See Passarelli (2020, pp. 88-9 and 96-97) for details. (d) In Italy and Estonia, compensation mandates are allocated to party and districts via an upper tier. (e) The nomination limit was raised to 4, for districts electing fewer than 4 seats, in 2021. (f) The representation threshold in Poland is 5% for single-party lists, but 8% for coalitions.

of each electoral system. As discussed, many institutional differences on either dimension are not explicitly modelled: the assumption, or rather the 'wager', is that actors will behave similarly in list-based systems that are sufficiently proportional and sufficiently preferential. The cases were mainly selected due to considerations of data availability; nonetheless, the sample contains a diverse range of PLPR institutional set-ups:

— Five OLPR systems (Cyprus, Finland, Poland, Italy, Peru) and four FLPR systems (Czechia, Slovakia, Belgium, Estonia).

— Three countries where voters *must* cast a preferential vote (Estonia, Finland, Poland) and six where they may do so, or otherwise cast only a list vote (Belgium, Cyprus, Czechia, Italy, Peru, Slovakia).

— Four countries where parties cannot nominate more candidates than the district magnitude (Belgium, Cyprus, Italy, Slovakia) and five where they can over-nominate (Czechia, Estonia, Finland, Peru, Poland).

— Three simple PR systems on the inter-party dimension (Belgium, Finland, Italy) and six where proportionality is corrected via the introduction of preference thresholds (Cyprus, Czechia, Estonia, Peru, Poland, Slovakia).

— Six countries employing a divisor formula (Belgium, Czechia, Estonia, Finland, Peru), and three using a quota formula (Cyprus, Italy, Slovakia).

— Six countries allowing multiple preference votes (Belgium, Cyprus, Czechia, Italy, Peru, Slovakia) and three restricting voters to one preference vote (Poland, Finland, Estonia).

Table 5.2: Median values of the variables of interest (maxima and minima in parentheses).

Dependent Variables

Country	Observations	v_1	N_c	v_s
Belgium	270	0.28 (0.12–0.62)	7.87 (2.38–14.99)	0.11 (0.02–0.59)
Cyprus	84	0.25 (0.11–0.65)	6.74 (2.07–15.10)	0.21 (0.04–0.59)
Czechia	241	0.16 (0.09–0.48)	13.18 (4.12–22.58)	0.11 (0.02–0.35)
Estonia	160	0.44 (0.17–0.94)	3.82 (1.13–8.22)	0.32 (0.03–0.85)
Finland	230	0.23 (0.09–0.9)	8.04 (1.22–23.09)	0.14 (0.02–0.90)
Italy	906	0.26 (0.07–0.96)	7.61 (1.09–27.31)	0.18 (0.01–0.96)
Peru	217	0.40 (0.15–0.75)	3.40 (1.69–15.54)	0.36 (0.01–0.75)
Poland	489	0.35 (0.09–0.90)	5.72 (1.23–18.26)	0.17 (0.01–0.73)
Slovakia	20	0.26 (0.17–0.39)	9.39 (5.82–13.02)	0.01 (0.001–0.02)

Independent Variables

Country	c	s	p	M	r
Belgium	16 (4–24)	2 (1–11)	16 (4–24)	15 (3–24)	1 (1–1)
Cyprus	11 (3–20)	1 (1–7)	3 (1–5)	11 (3–20)	1 (1–1)
Czechia	22 (14–36)	2 (1–11)	4 (4–4)	12 (5–26)	1.8 (1.31–2.8)
Estonia	10 (7–17)	2 (1–6)	1 (1–1)	8 (5–15)	1.2 (1.13–1.4)
Finland	17 (2–36)	2 (1–11)	1 (1–1)	17 (6–36)	1 (1–2.33)
Italy	20 (3–54)	2 (1–20)	4 (3–4)	21 (2–55)	1 (1–1)
Peru	5 (3–36)	1 (1–15)	2 (2–2)	5 (1–36)	1 (1–4)
Poland	23 (13–40)	2 (1–12)	1 (1–1)	12 (7–20)	2 (2–2)
Slovakia	150 (148–150)	15.5 (10–83)	4 (4–4)	150 (150–150)	1 (1–1)

Turning now to the observed values of the variables of interest, table 5.2 reports median, maxima and

minima for the dependent (v_1, N_c, v_s) and independent (c, s, p, r, M) variables, computed at list level and broken down by country. A descriptive analysis of list-level variables is also reassuring with regards to the assumption that parties always over-nominate: the mean ratio of c/M across the whole sample is 1.30, while the mean value of r is 1.31, meaning that the institutional limit to nominations is effectively tantamount to the number of candidates nominated by seat-winning parties. Even in Poland, where r is highest as lists may nominate up to $2M$ candidates (and in theory they may field as few as $M/2$), on average seat-winning lists nominate a number of candidates that is 1.97 the district magnitude.

5.5 Methodology

Restating the conclusions of the theoretical section, I derived six quantitative predictions in the form $Y = X^k$ linking products of electoral quantities to indicators of intra-party competitiveness: one for each of the three dependent variable in the list-level model, and one for each of the three dependent variable in the district-level institutions-only model. These are effectively our hypotheses: given a product of electoral system quantities as the base X of the function, the exponent k is expected to take a certain value derived theoretically (in practice, $-\frac{1}{4}$ or $\frac{3}{8}$). I notate the value of the exponent expected from theory as $\hat{k}$ and the value obtained from regression as β in the rest of the analysis. Summing up:

Y variable	X (list-level)	X (district-level)	$\hat{k}$ (expected slope)
v_1 (first candidate's share)	scp	$prM^{\frac{11}{8}}$	$-\frac{1}{4}$ or -0.25
N_c (eff. number of candidates)	scp	$prM^{\frac{11}{8}}$	$\frac{3}{8}$ or 0.375
v_s (last eligible cand. share)	s^4cp	$prM^{\frac{5}{2}}$	$-\frac{1}{4}$ or -0.25

The empirical section proceeds in two steps:

— In subsection 5.6.1, I employ regression analysis to test the bias of the slope predictions $\hat{k}$ in the table above. Moreover, for v_1 and v_s, I compare the revised intra-party models' performance to their equivalents in SBW. The bias of the models is measured here as the discrepancy between the expected and observed slopes, normalised by the standard error.

— In subsection 5.6.2, I compare the precision of the individual predictions of the models of intra-party quantities with the better established predictions of their inter-party analogues of the seat-product model (SPM).[18] Specifically, the models for the fractional share of first-ranked candidates in a district (v_1) are compared with the SPM predictions for the fractional share of seats for the largest party in an assembly (σ_1); the models for the effective number of candidates (N_c) are compared with the SPM predictions for the effective number of parliamentary parties (N_S). The precision of the predictions is measured by computing the deviation-from-prediction of each observation, using the d index proposed by Nemčok and Šedo (2018) and explained shortly in subsection 5.5.2. The d index is computed for (1) the values of v_1 and N_c of each list compared to the list-level model's predictions, (2) the median

[18] Bias and precision are here used to denote different aspects of model performance. Bias indicates error in *the extent to which a model describes the overall relationship between variables in the sample*, while precision indicates *the extent to which the model can predict individual observations in the sample*. The latter is therefore a summary measure of bias and variance, and applies to observations rather than samples.

values of v_1 and N_c for each district compared with the district-level model's predictions, and (3) the values of σ_1 and N_S for each assembly election compared with the SPM's predictions.

5.5.1 Model Bias and Comparison with the SBW Model

In the first empirical section, I present graphical summaries of each of the six hypotheses, where the predicted and observed functional forms of the regressions are plotted against the data. Moreover, I report the exponential slope coefficients obtained by performing regression analysis on the full sample, only on the sample of elections contested under OLPR rules, and only on the sample of elections contested under FLPR rules. As a measure of coefficient bias, I compute for each of these the absolute difference between the observed slope β and predicted estimate $\hat{k}$ normalised by the standard error of the estimate (standard errors are clustered at the election-district level).

To test the models, I follow the methodological recommendations in Taagepera (2008) and employ fixed-intercept exponential regressions where the relationship between variables is modelled as $Y = X^{\beta}$.[19] This is equivalent to a log-log regression where the intercept is set to be zero, so that fitted values may not exceed $X^0 = 1$. Regression parameters are thus constrained to predict positive values of the dependent variable, lying between 0 and 1 when k is negative and larger than 1 when k is positive. Moreover, the fixed-intercept serves an 'anchor point' (Taagepera, 2008, pp. 44-45) that prevents us from making illogical prediction: we *know* a priori that in the limit where a PLPR contest collapses into a single-member district plurality race, $c = s = p = scp = 1$ and all three quantities v_1, N_c and v_s will be 1. In other words, the seat-winning party will field only one candidate, who will come 'both first and last' with 100% of the votes. Deriving an empirical intercept other than 1 would predict an absurdity.

In the presentation of the results for v_1 and v_s, I also compare the model performance against the existing predictions for these quantities in SBW.[20] Their estimation of v_1 in open-list systems, as discussed, is

$$v_1 = c^{-\frac{1}{2}} \tag{5.11}$$

As for v_s, the SBW model's expectation is that in an open-list PR system the share of preference votes of the last elected candidate is comprised between v_1 and the last elected candidate's share under SNTV rules, which is separately estimated as c^{-1}. Hence, in principle,

[19] The estimator used is simple OLS. In the appendix section H.3, I present the results of the same tests employing 'symmetric regression', a different estimator favoured by Taagepera (2008). Moreover, I present slope estimates dropping one country from the sample, to show that the predictive accuracy is not primarily driven by any one institutional set-up (section H.4).

[20] To my knowledge, there are no existing logical models for N_c.

$$v_s = (c^{-\frac{1}{2}} \times c^{-1})^{\frac{1}{2}} = c^{-\frac{3}{4}} \tag{5.12}$$

However, SBW prefer to employ the *observed* value of the exponent for v_1 instead of the 'ignorance-based' value of $-\frac{1}{2}$, so that after observing the empirical value of the slope β, where $v_1 = c^{\beta}$, v_s is estimated as

$$v_s = (c^{\beta} \times c^{-1})^{\frac{1}{2}} = c^{\frac{-1+\beta}{2}} \tag{5.13}$$

Both versions – which I call respectively the 'uncorrected' and 'corrected' SBW models – are computed and compared to my model's predictions for v_1. The extent to which the observed relationship between c and v_s deviates from these predictions is again calculated by taking the absolute value of the difference divided by the standard error.

5.5.2 Model Precision Compared to Inter-Party Models

The final empirical section addresses the question posed at the very beginning of the paper: *to what extent can we make predictions about intra-party competition in the same way as the seat-product model does with respect to inter-party competition?* To do so, I compare the list- and district-levels models for v_1 and N_c with their inter-party analogues:[21] respectively, the fractional share of seats of the largest party (σ_1) and the effective number of parties in an assembly (N_S). As noted at the start of the paper, the inter-party quantity predictions of the seat-product model consist of functions of the product of district magnitude (M) and assembly size (S). The formulas for σ_1 and N_S given in Taagepera (2007) and Shugart and Taagepera (2017) are, respectively:

$$\sigma_1 = (MS)^{-\frac{1}{8}} \tag{5.14}$$

$$N_S = (MS)^{\frac{1}{6}} \tag{5.15}$$

Following Nemčok and Šedo (2018), in this part of the analysis I compute for each observation a measure of discrepancy d as $\log_{10}(\frac{y}{\hat{y}})$, where y is the observed value and $\hat{y}$ is the prediction. The logarithmic function of the ratio between observed and predicted values takes the value of 0 when the observation perfectly mirrors

[21] There is no obvious inter-party counterpart to v_s. An analysis of deviation-from-prediction of the v_s model analogous to the one conducted for v_1 and N_c is presented in the appendix, without comparison to inter-party quantities.

theoretical expectations, positive values when the observed quantity is larger than expected and negative values when it is smaller. More specifically, d expresses the factor by which the prediction is off: if the observed value is twice the predicted value, d will be $\log_{10}(2) \approx 0.3$; if it is half the predicted value, d will be $\log_{10}(0.5) \approx -0.3$. Again following Nemčok and Šedo (2018), I use these two values as the bounds of 'tolerable' deviation, and compute the percentage of cases that fall in between, and thus are predictable from input variables within a factor of 2. I also report measures of central tendency for $|d|$, from which it is possible to extrapolate the average factor of error of the model predictions, as $(10^{|d|} - 1) \times 100\%$. For instance, predictions that are twice or half the observed value will both have $|d| \approx 0.3$, and the factor of error will be $(10^{0.3} - 1) \times 100\% = (2 - 1) \times 100\% = 100\%$.

For the intra-party model, I present two sets of results: first, I compute d for v_1 and N_c for each of the 2,617 list-in-district observations in the intra-party dataset described in section 5.4, using the predictions of the list-level model. Then, I compute d for the *median values* of v_1 and N_c in each of the 549 districts in my data, using the predictions of the district-level model.[22] As discussed, the district-level predictions will take the same value for each list, and they are meant to capture the level of competitiveness for any list in expectation. For the inter-party analysis, I replicate and extend the analysis in Nemčok and Šedo (2018), using the country-election level data they gathered on institutional characteristics and election results in 560 democratic elections.[23]

Therefore, I compute the index d of six quantities: the list-level observed intra-party quantities v_1 and N_c; the district-level median intra-party quantities $\tilde{v}_1$ and $\tilde{N}_c$; and the election-level inter-party quantities σ_1 and N_S. By the definition of d, the formulas are as follows:

Intra-Party Quantity (List-Level)	d index	Intra-Party Quantity (District-Level)	d index	Inter-Party Quantity (Election-Level)	d index
v_1	$\log_{10}\left(\frac{v_1}{(scp)^{-\frac{1}{4}}}\right)$	$\tilde{v}_1$	$\log_{10}\left(\frac{\tilde{v}_1}{(prM^{\frac{11}{8}})^{-\frac{1}{4}}}\right)$	σ_1	$\log_{10}\left(\frac{\sigma_1}{(MS)^{-\frac{1}{8}}}\right)$
N_c	$\log_{10}\left(\frac{v_1}{(scp)^{\frac{3}{8}}}\right)$	$\tilde{N}_c$	$\log_{10}\left(\frac{\tilde{N}_c}{(prM^{\frac{11}{8}})^{\frac{3}{8}}}\right)$	N_S	$\log_{10}\left(\frac{N_S}{(MS)^{\frac{1}{6}}}\right)$

[22] In section 5.6.2, I use the median values, as logical models are meant to yield predictions that are equally likely to be above or below the real values. In section K of the Appendix, I repeat this exercise with the mean values of the variables. Medians of v_1 and N_c are notated as $\tilde{v}_1$ and $\tilde{N}_c$, while means are notated as $\bar{v}_1$ and $\bar{N}_c$.

[23] The Nemčok-Šedo dataset of electoral quantities is described in the Appendix, section I.

5.6 Empirical Tests

5.6.1 Model Bias and Comparison with the SBW Model

First-ranked Candidate Share (v_1)

Regressing the observed values of v_1 on *scp* returns the fixed-intercept exponential function of $Y = X^{-0.262}$ with the coefficient having a 95% confidence interval $(-0.254, -0.270)$. The list-level prediction $\hat{k} = -0.25$ therefore falls just outside the confidence interval, narrowly over-predicting first-candidate share. As for institutions-only model, the regression of v_1 on the product $prM^{\frac{11}{8}}$ returns the fixed-intercept exponential function $Y = X^{-0.253}$ with the coefficient having a 95% confidence interval (-0.260, -0.246), therefore including the prediction $\hat{k} = -0.25$.

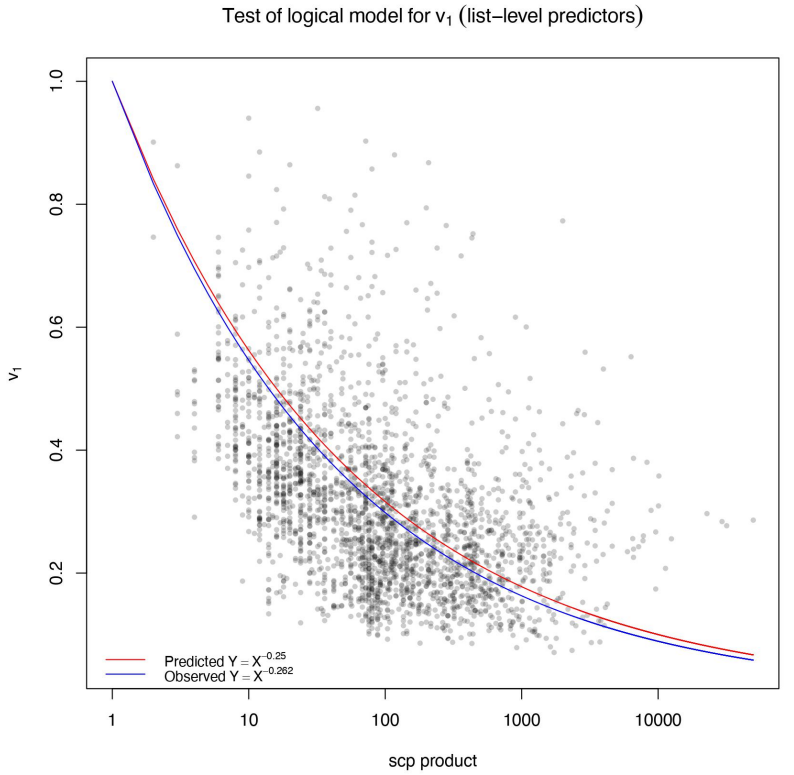
Figure 5.4 plots the predicted and observed exponential functions for the two models. In a pattern that we will observe across model tests of v_1 and N_c, the Slovak lists on the far right of the plots, which take the highest values of the product as they refer to parties competing in a district of magnitude 150, have substantially higher values than predicted. While the institutions-only prediction presents even less bias than the already rather accurate list-level model, it is clear from the plot there is much more scatter around it, as the independent variable takes the same value for all lists-in-district regardless of their actual inter-party competitiveness.

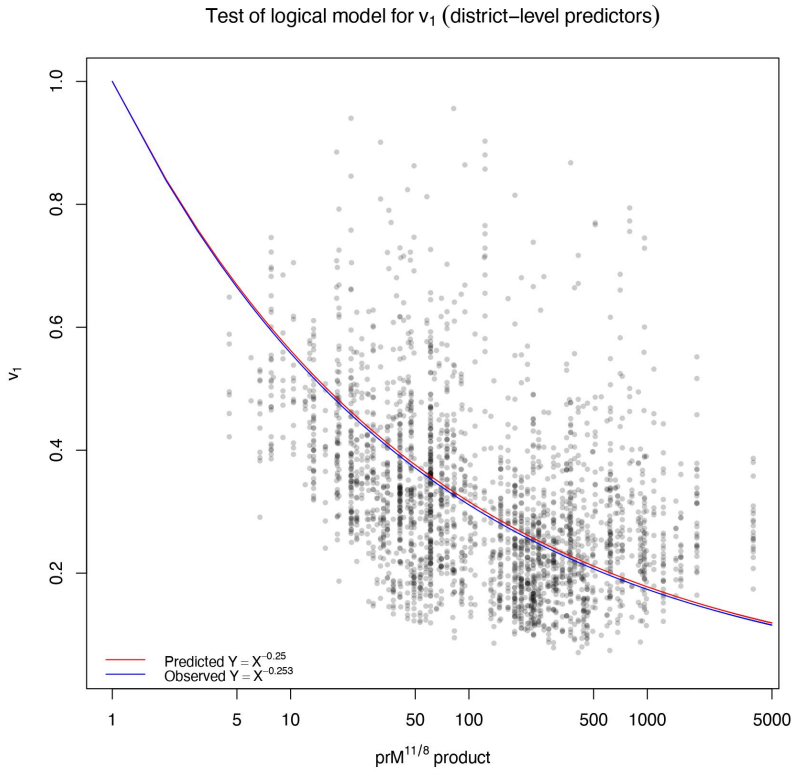
Table 5.3: Comparison of model fits for predictive models of the first-ranked candidate's fractional share of preference votes (v_1).

		Full Sample		Open List Only		Flexible List Only	
	$\hat{k}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$
List-Level Model	-0.250	-0.263	3.088	-0.272	5.753	-0.243	0.806
District-Level Model	-0.250	-0.253	0.992	-0.259	2.264	-0.243	1.018
SBW Model	-0.500	-0.435	11.574	-0.423	13.456	-0.468	2.270

Table 5.3 reports a comparison of the slope coefficients obtained from regressing v_1 on the products of list-level and district-level quantities, against the performance of the SBW model, where v_1 is regressed on c. Absolute values of the discrepancy between predicted and observed values of the coefficients normalised by the standard error are reported across the full, open-list and flexible-list samples. While both the refined models outperform the coarse model, the district-level model's predictions are remarkably accurate, with $\hat{k}$ falling within one or two standard errors from the prediction, depending on the sample specification. In the full sample, the SBW's slope falls over 11 standard errors away from the observed slope: a much larger bias

Figure 5.4: Test of list- and district-level predictions for the first-ranked candidate's share of preference votes: predicted and observed slope of v_1 regressed on scp and $prM^{\frac{11}{8}}$.





than the 3 standard errors of the list-level 'revised' model and the 1 standard error bias of the district-level 'revised model'. The SBW model for v_1 performs 'best' in the flexible-list sub-sample: this is very surprising as it was meant for and tested on open-list systems only.

Effective Number of Candidates (N_c)

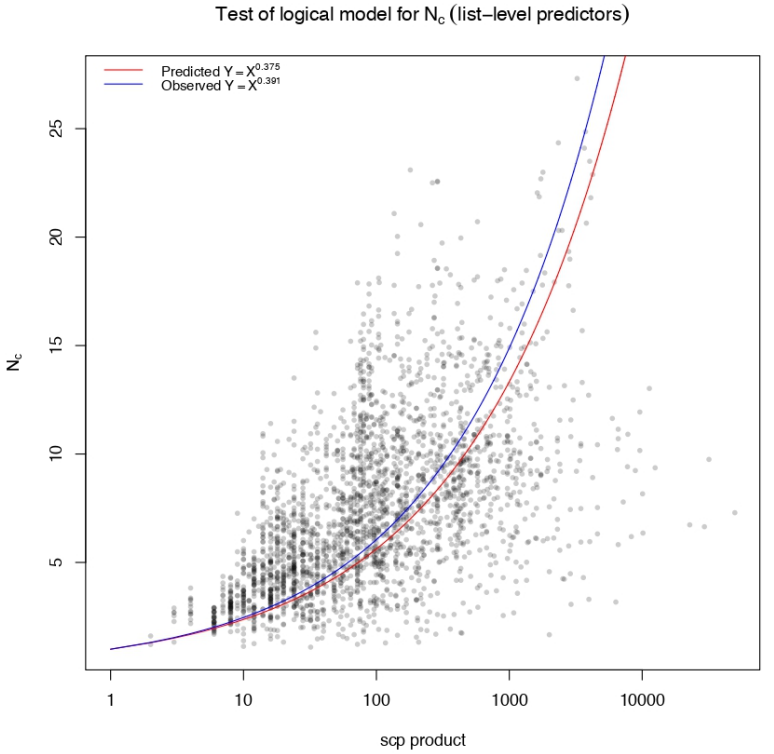
Regressing the observed distribution of the effective number of candidates N_c on *scp* returns a fixed-intercept exponential function of $Y = X^{0.391}$ with the coefficient having a 95% confidence interval $(0.381, 0.40)$. Again, the prediction $\hat{k} = 0.375$ comes close to the observed value of the slope but somewhat understates intra-party competitiveness, in this case by narrowly under-predicting N_c. Regressing N_c on $prM^{\frac{11}{8}}$ returns a fixed-intercept exponential function of $Y = X^{0.379}$ with the coefficient having a 95% confidence interval (0.367, 0.390), therefore including the predicted value of $k = 0.375$. Figure 5.5 plots the predicted and observed exponential functions for the two models, and table 5.4 reports the observed slope coefficients across different specifications of the sample. As in the case of v_1, the institutions-only model is noticeably less biased than the list-level model and its performance presents little variance due to sample specification. Again, the graphs show clear outliers occurring when the *scp* product and its institution-only equivalent $prM^{\frac{11}{8}}$ take the highest observed values. These occur in Slovakia's nationwide district.[24]

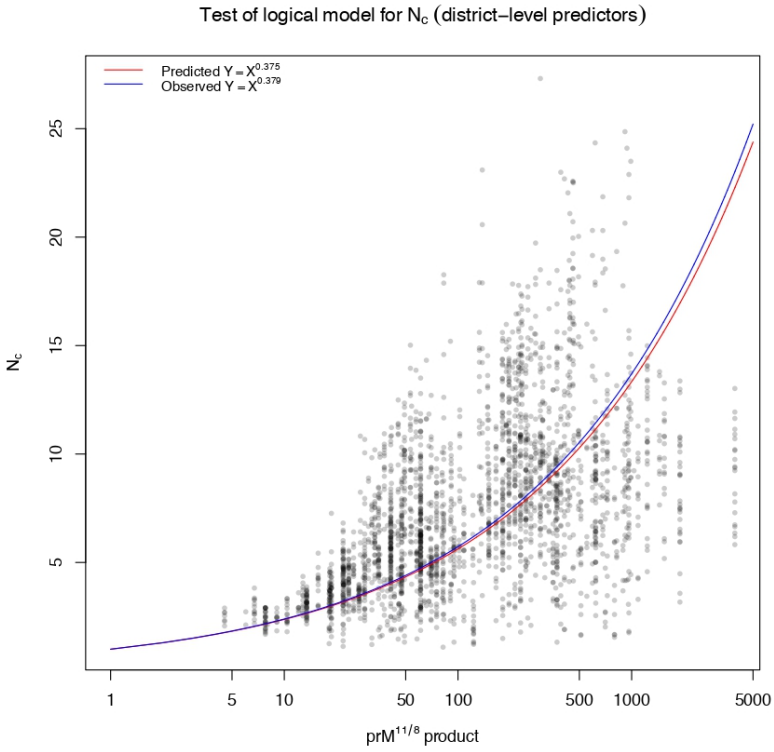
Table 5.4: Model fits for predictive models of the effective number of candidates.

		Full Sample		Open List Only		Flexible List Only	
	$\hat{k}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$
List-Level Model	0.375	0.391	3.136	0.402	5.414	0.369	0.582
District-Level Model	0.375	0.379	0.888	0.385	1.878	0.367	0.928

[24]For further discussion of these outliers, see section 5.7.3.

Figure 5.5: Test of list- and district-level predictions for the effective number of candidates: predicted and observed slope of N_c regressed on scp and $prM^{\frac{11}{8}}$.





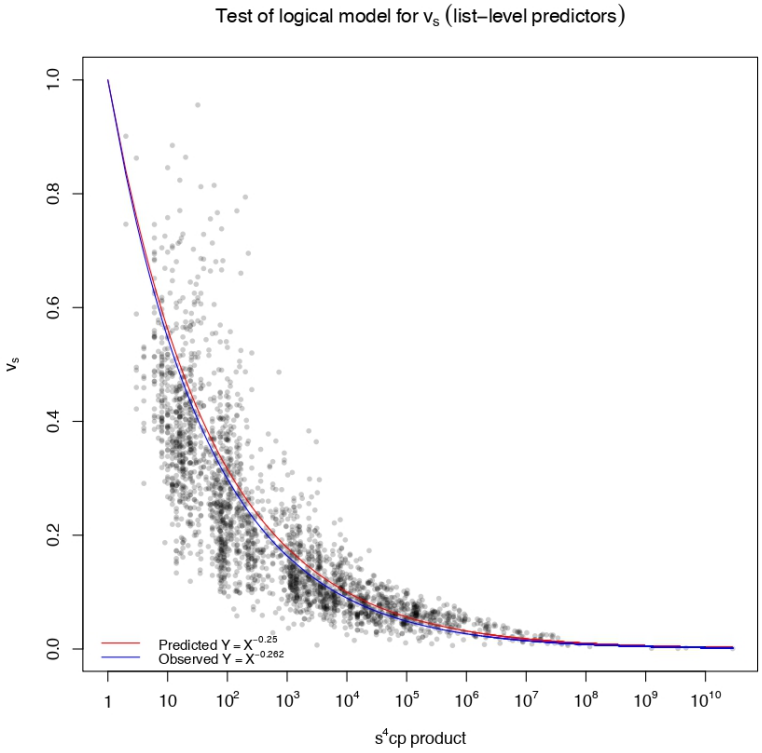
Last Eligible Candidate's Share of Preferential Votes

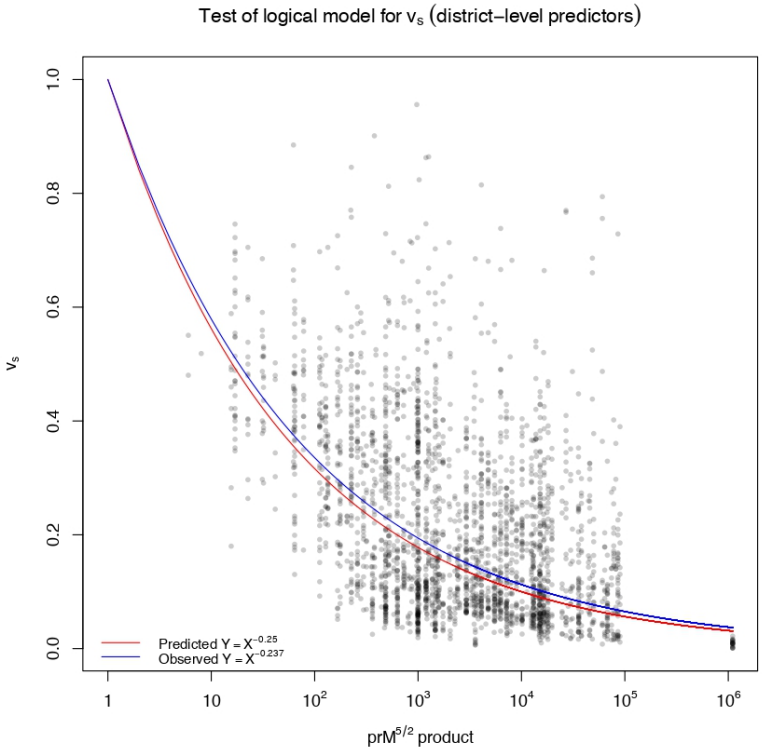
Regressing v_s on s^4cp returns a fixed-intercept exponential function of $Y = X^{-0.262}$ with the coefficient having a 95% confidence interval (-0.266, -0.258). Regressing v_s on $prM^{\frac{5}{2}}$ returns a fixed-intercept exponential function of $Y = X^{-0.237}$ with the coefficient having a 95% confidence interval (-0.244, -0.230). As shown in figure 5.6, in this case we find that the list-level model somewhat under-predicts v_s, while the district-level model somewhat over-predicts it. Table 5.5 reports coefficients across specifications of the sample for the list- and district-level models, alongside the normalised discrepancy from the expected value of $\hat{k}$. In this case, the SBW model – especially in its 'corrected' version where $\hat{k}$ depends on the empirical slope of v_1 as a function of c – comes close to the model performance of the two 'refined' models. Indeed, in the sub-sample of flexible-list observations, it outperforms them. However, there is a large amount of variation in performance of the SBW model fit across specifications of the sample: something that we do not observe to the same extent for the 'refined' list- and district-level models. In the full sample of observations, the list- and district-level models for v_s present more bias than in the models for v_1 and N_c described above, but nonetheless they still both outperform the SBW model.

Table 5.5: Comparison of model fits for predictive models of the last eligible candidate's preference share.

		Full Sample		Open List Only		Flexible List Only	
	$\hat{k}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	β	$\frac{\lvert\beta-\hat{k}\rvert}{se}$
List-Level Model	-0.250	-0.262	6.928	-0.259	4.206	-0.270	6.971
District-Level Model	-0.250	-0.237	4.636	-0.229	7.800	-0.256	1.130
SBW Model (uncorrected)	-0.750	-0.655	13.572	-0.622	24.492	-0.750	0.018
SBW Model (corrected)	-0.717	-0.655	8.897				
	-0.712			-0.622	12.798		
	-0.734					-0.750	1.008

Figure 5.6: Test of list- and district-level predictions for the last eligible candidate's share of preferences: predicted and observed slope of v_s regressed on s^4cp and $prM^{\frac{5}{2}}$.





5.6.2 Model Precision and Comparison with Inter-Party Models

First-ranked candidate (v_1) and largest party (σ_1) predictions compared

Figure 5.7 plots on the y-axis the values of the discrepancy index d – which correspond to the log-transformed ratio of observed and predicted values – and on the x-axis is the base product of each of the three predictive models under examination. Specifically, the top panel visualises the logged ratio of observed values of v_1 and the predicted quantity $(scp)^{-\frac{1}{4}}$ for all seat-winning lists in the dataset described in section 5.4. The middle panel represents the logged ratio of observed median values of v_1 in each district and the prediction $(prM^{\frac{11}{8}})^{-\frac{1}{4}}$, drawing on the same data. The bottom panel visualises the logged ratio of observed values of σ_1 (fractional share of the largest party in an assembly) and the SPM predicted values of this quantity, computed as $(MS)^{-\frac{1}{8}}$ on the Nemčok-Šedo dataset of elections.

It is evident from the plots that list-level predictions of v_1 tend to fall farther from zero than those of the other models. However, when it comes to predicting *median* intra-party competition, the panel plot for the district-level model is visually very similar to that of the SPM's predictions. Not only do the vast majority of the values of d fall within -0.3 and 0.3, indicating that a prediction is within a factor of 2 from the observed value, but the shapes of the distributions of d in the second and third panels are also quite similar, even though they represent distinct quantities computed on distinct datasets. Specifically, both predictions of first-ranked candidate shares and of largest party shares tend to 'miss the mark' most clearly when the models' base products take the highest values – i.e. for those sets of electoral institutions that in theory should be most conducive to competition – and in both cases the prediction overstates competitiveness relative to reality. This suggests that there are some upper constraints of political nature to the fragmentation of a system can handle when it is least constrained: under highly permissive rules 'on paper', party and preference votes will tend to concentrate in ways that cannot be accounted for simply by institutional factors (or at least not the parsimonious set of institutional factors that are sufficient to predict competition in other contexts).[25]

Table 5.6: Summary indicators of deviation from prediction: models for v_1, $\tilde{v}_1$ and σ_1 compared

	median of $\lvert d \rvert$		mean of $\lvert d \rvert$		share $d < log_{10}(2)$
	value $\lvert d \rvert$	% error	value $\lvert d \rvert$	% error	and $d > log_{10}(0.5)$
list-level model (v_1)	0.119	31.6%	0.149	40.9%	87.9%
district-level model ($\tilde{v}_1$)	0.083	21.1%	0.103	26.8%	97.1%
seat-product model (σ_1)	0.091	23.4%	0.105	27.4%	96.2%

[25]Interestingly, Slovakia is a case of both phenomena occurring at the same time: a highly permissive nationwide PR system that should have relatively small largest parties and low first-ranked candidate shares, but presents moderate levels of vote concentration on both dimensions, at least in some elections.

Figure 5.7: Comparison of deviation from prediction for v_1 (list-level model), median $\tilde{v}_1$ (district-level model) and σ_1 (seat-product model). Dashed lines represent values of d corresponding to values where the observed value is either twice or half the prediction.

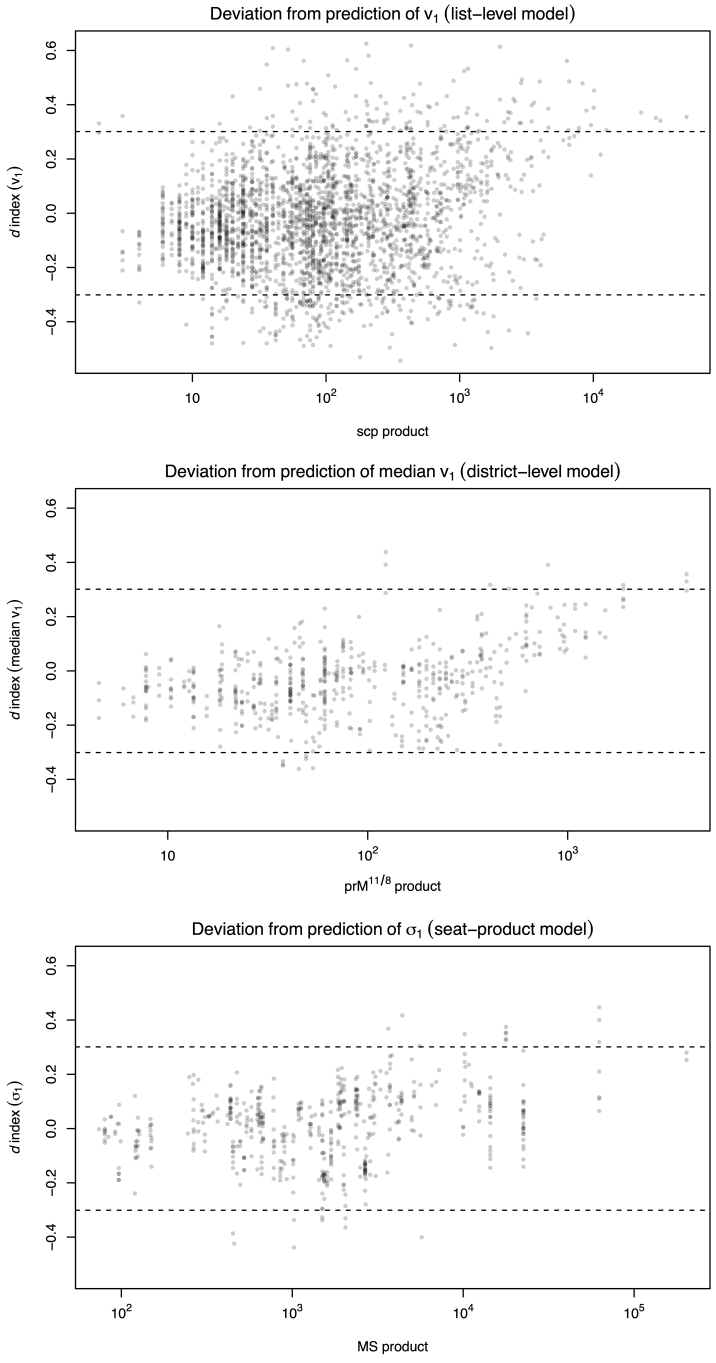


Table 5.6 confirms the conclusions drawn from the visual presentation of the data. The average absolute values of d are highest for the list-level model of v_1, and similarly low for the district-level model of $\tilde{v}_1$

and the seat-product model for σ_1. Alongside the measures of central tendency, the table also reports the associated error factors: the median (mean) prediction of the list-level model is off by 32% (41%); the median (mean) prediction of the district-level model is off by 21% (27%); the median (mean) prediction of the seat product model is off by 23% (27%). The logical models under consideration can predict the largest party's fractional share of seats and the median fractional share of first candidate's preference votes from institutional quantities for almost all observations (96–97%) in the samples within a factor of 2. However, list-specific predictions are clearly less precise, with 12% of the model expectations being more than double or less than half the actual values.

Effective Number of Candidates (N_c) and Parties (N_S) Predictions Compared

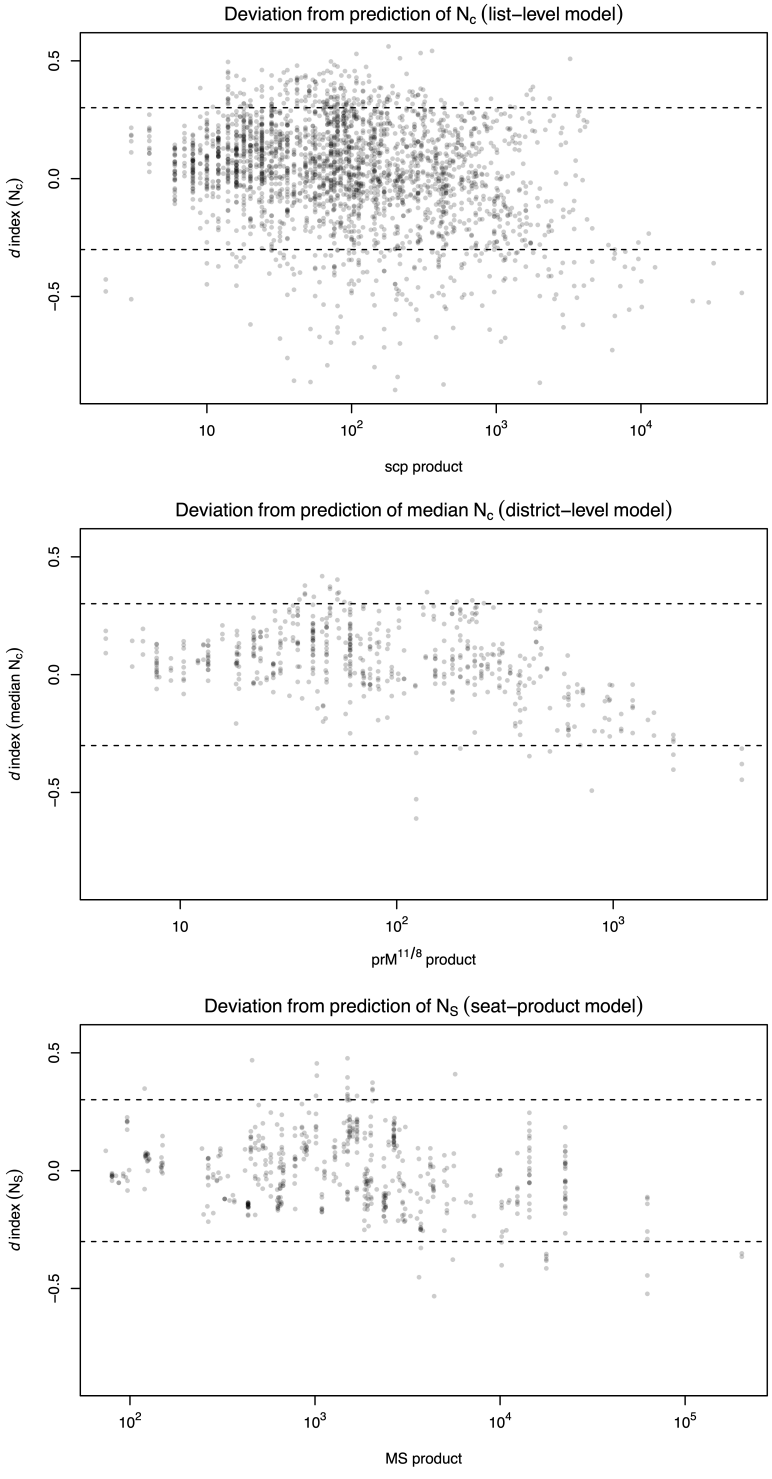
This subsection reproduces the same analysis of 'deviation from prediction' but for a different set of dependent variables: the effective number of candidates of a list (N_c), the median effective number of candidates for a list in a district ($\tilde{N}_c$), and the effective number of parties in an assembly (N_S).

Figure 5.8 plots the values of d against the models' base products. In keeping with the observations of the previous subsection, the visualisations show that the predictions of the list-level model are much more widely scattered, and hence less precise, than those of the district-level model and the SPM. And once again we find that the distribution of the deviation-from-prediction values for median effective number of candidates in a district and the effective number of parties in an assembly are strikingly similar. In both cases, the vast majority (94.5%) of values lie within the bounds of 'tolerable' discrepancy comprised between -0.3 ($\log_{10} 0.5$) and 0.3 ($\log_{10} 2$); and in both cases, the most notable deviations are instances of *over-prediction* of competitiveness observed where the base products take the highest values observed. Table 5.7 reports measures of central tendency for the distribution of d across models, which confirm that the district-level model and the seat-product model are about as precise in predicting the effective number of parties/candidates in their respective samples.

Table 5.7: Summary indicators of deviation from prediction: models for N_c, $\tilde{N}_c$ and N_S compared

	median of $\lvert d \rvert$		mean of $\lvert d \rvert$		share $d < log_{10}(2)$
	value $\lvert d \rvert$	% error	value $\lvert d \rvert$	% error	and $d > log_{10}(0.5)$
list-level model (N_c)	0.142	38.6%	0.169	47.6%	84.3%
district-level model ($\tilde{N}_c$)	0.106	27.8%	0.126	33.8%	94.5%
seat-product model (N_S)	0.113	29.6%	0.124	33%	94.5%

Figure 5.8: Comparison of deviation from prediction for N_c (list-level model), median $\tilde{N}_c$ (district-level model) and N_S (seat-product model) . Dashed lines represent values of d corresponding to values where the observed value is either twice or half the prediction.



Summing up, the answer to the guiding question of whether we can predict intra-party quantities from institutional variables in the same way as we can predict inter-party is thus arguably both a 'no' and a 'yes'. On the one hand, the list-level model presented fails to estimate first-ranked candidate's preference share *for individual lists* with the same precision of the SPM's prediction for the fractional share of seats of the largest party in an assembly. On the other hand, at district level, we can derive a prediction of the median value of v_1 and N_c for seat-winning parties from purely institutional variables that is just as precise as the SPM's prediction for σ_1 and N_S. Of course, one may retort that even in this case the terms of the comparisons presented are not entirely 'fair': although the number of districts used to test district-level predictions (549) and the number of elections used to test the SPM predictions (560) are almost identical, the SPM draws on data from 40 countries, while the intra-party model only on nine. At the same time, to compare 'like with like', we ideally may want to compare *district-level* inter-party predictions with *district-level* intra-party prediction. These are both valid caveats to the comparison presented, which will be addressed in future research through further data collection of district-level and list-level results.

5.7 Discussion and Conclusion

Intra-party competition has often been thought of as unpredictable and idiosyncratic in comparison to inter-party competition, to the extent that most works on PLPR begin by commenting on how poorly understood and understudied this aspect of electoral institutions are.[26] This paper aims to join a small but growing group of theoretical (Shugart, Bergman and Watt, 2013; Buisseret and Prato, 2020; Buisseret et al., 2022) and empirical (Selb and Lutz, 2015; Renwick and Pilet, 2016; Blom-Hansen et al., 2016; Däubler and Hix, 2018; Passarelli, 2020; Cheibub and Sin, 2020; Dodeigne and Pilet, 2021) works aiming to identify order in the apparent chaos of intra-party competition in candidate-centred electoral systems. To conclude the paper, I present a possible application of the framework outlined to political practice, discuss the implications of my analysis for further study of PLPR systems, and note some shortcomings of the models.

5.7.1 An Application to Institutional Choice

What practical use is it to know how preference votes 'should', in expectation, be distributed within a PLPR list? As an illustration of the potential relevance of the models presented to institutional design, let us consider how an institutional engineer may decide to fix the preference threshold in a FLPR system, so as to achieve a desired balance-of-power between voters and parties. Consider a FLPR system where a candidate must receive at least a fractional share t of preference votes to be elected on the basis of preferential votes; if the number of seats allocated to a list exceeds the number of candidates that meet this threshold, the remaining seats are allocated according to the ballot position determined by the party. Thus, t effectively indicates how much control parties and voters have on the intra-party allocation of seats: the higher the fractional value of t, the more candidates will be elected from their ballot position; the lower the fractional value of t, the more candidates will be elected from preference votes.

From the theoretical considerations above, we can identify three key values of t expressed as a fractional share of preference votes: (1) the value of t that is equally likely to produce one or no candidates elected on preferences in a district (t_1), (2) the value of t that is equally likely to produce M or $M-1$ candidates elected on preferences in a district (t_M); and (3) the value of t whereby in expectation 50% of MPs are elected on preference vote ranking and 50% of MPs are elected due to ballot position ($t_{\frac{M}{2}}$). t_1 is realised when the threshold equals the expected fractional share of preference votes of the relative top-performing seat-eligible

[26] For instance, for Buisseret and Prato (2018, p.1) "[i]n spite of their widespread use, open- and flexible-list proportional representation [...] systems have received little attention from empirical scholars and almost none from theoretical scholars. In large part, this is due to the fact that these systems vary tremendously in their operation across countries, bedeviling attempts at classification and limiting scholarly efforts to the analysis of specific cases."

candidate. As, in expectation, such candidate is the first-ranked candidate (who gets a preference share of v_1) of the smallest party (which expects to win one seat), it follows that

$$t_1 = (p \times rM \times 1)^{-\frac{1}{4}} = (prM)^{-\frac{1}{4}} \tag{5.16}$$

As for t_M, the institutional engineer would set the threshold to equal the expected preference share of the seat-eligible candidate gaining the smallest relative preference share in the district. As, in expectation, such candidate is the last-ranked candidate (who gets a preference share of v_s) of the largest party (gaining $M^{\frac{3}{4}}$ seats), it follows that

$$t_M = \left(p \times rM \times (M^{\frac{3}{4}})^4\right)^{-\frac{1}{4}} = (prM^4)^{-\frac{1}{4}} \tag{5.17}$$

I conjecture that $t_{\frac{M}{2}}$ may be approximated as the geometric mean of t_1 and t_M, so that

$$t_{\frac{M}{2}} = \left((prM)^{-\frac{1}{4}} \times (prM^4)^{-\frac{1}{4}}\right)^{\frac{1}{2}} = (prM^{\frac{5}{2}})^{-\frac{1}{4}} \tag{5.18}$$

Which version of t should the institutional engineer choose? t_1 is of no use: a candidate popular enough to get that high a share of preference votes is likely to be already in a seat-eligible position. Effectively, t_1 produces what Däubler and Hix (2018) term 'weakly flexible' lists, where it is unlikely that any candidate beyond the top-ranked gather enough preference votes to be elected on the basis of their personal support. t_M, conversely, would be an apt choice if the institutional engineer wanted to have a permissive 'quasi-open' list system, which only prevents candidates with very little personal appeal from lucking into a parliamentary seat on a sparse number of preference votes. Under t_M, we are guaranteed to observe what Däubler and Hix (2018) term 'strongly flexible' lists. In fact, small parties are effectively competing under open-list rules, because the threshold is too low to ever make a difference, while large parties retain occasionally some degree of control over the intra-party allocation but only for the last few seats. The choice of $t_{\frac{M}{2}}$ would be a compromise between the two. On the one hand, parties retain a substantial deal of control over election outcomes: though half of the candidates in a district are elected on preferences, most of them would presumably be in seat-eligible ballot positions anyway. However, such a system is still flexible enough to reward – occasionally – strong performances from down-ballot candidates, making preferential voting meaningful. Table 5.8 reports the share of candidates who would have been elected on preference votes under the different district-level specifications of the threshold derived above, *if these elections in the sample*

had be conducted under such FLPR rules. Albeit rudimentary, this analysis is encouraging: in no case the thresholds are so extreme as to become trivial (some, but not all, candidates are elected on preferences in all cases) and $t_{\frac{M}{2}}$ is on average quite close to the speculative prior that it might produce a 50-50 split between candidates elected on preference ranking and candidates elected on ballot position. However, cross-country variation is substantial.

Table 5.8: Simulation of PLPR outcomes under different values of hypothetical preference thresholds.

	Share of candidates elected on preferences under		
Country/Sample	t_1	$t_{\frac{M}{2}}$	t_M
Belgium	0.18	0.65	0.98
Cyprus	0.02	0.57	0.99
Czechia	0.02	0.39	0.95
Estonia	0.10	0.58	0.84
Finland	0.03	0.36	0.97
Italy	0.08	0.49	0.99
Peru	0.05	0.55	0.92
Poland	0.06	0.39	0.75
Slovakia	0.04	0.27	0.65
Country Average	0.07	0.47	0.89
Full Sample	0.07	0.46	0.92

Of course, in most existing FLPR systems, thresholds are set as a function of *list* rather than *preference* votes, so that the values of t derived above are directly applicable to real-world electoral systems only in cases where $p = 1$ and preferential vote is mandatory.[27] If $p > 1$, preference votes exceed list votes, and thus t would have to be adjusted upwards by a factor of $p^{\frac{1}{2}}$, the expected number of preferences-per-voter. If preferential voting is optional, then list votes exceed preference votes, and t would have to be adjusted downwards by the expected fractional share of voters who cast a preference votes. This is a less readily predictable factor, not just because it may largely depend on political rather than institutional factors (party-voter linkages, voter engagement, democratic experience, party system and personnel stability etc.), but also because the extent to which voters will make use of a preference vote is likely endogenous to list flexibility itself (Däubler, 2020; Renwick and Pilet, 2016, pp. 217–248).

5.7.2 Implications for the Study of PLPR

The most novel aspect of the model presented is the introduction of proxies for expectations of inter-party performance as an input variable, in the form of s (the number of seats effectively gained by a list) for the

[27] An even less immediately tractable case is when t is expressed as a function of the electoral quota. However, this case could be tackled drawing on recent advances in modelling party *vote* share (Shugart and Taagepera, 2017, pp. 125-138), and thereby expressing expected preference shares as fractions of expected quotas attributed to each party.

list-level model, and $M^{\frac{3}{8}}$ (the expected number of seats for the median list) for the district-level model. The role of actors' expectations in shaping electoral system outcomes has long been studied as a key mechanism shaping inter-party competition. This is effectively what Duverger (1959) termed the 'psychological effect' of district magnitude, and scholars working in his tradition expanded to encompass parties', as well as voters', strategic behaviour (Cox, 1997). Yet, the recognition of a similar mechanism being at play at the intra-party level departs from the standard theoretical assumption, for instance made by Shugart, Bergman and Watt (2013), that district magnitude affects list competitiveness only insofar as it poses constaints on parties' nominating behaviour. The revised model, conversely, takes into account two avenues through which M is connected to intra-party competition: by constraining candidate nominations *and* by providing candidates (and voters) with priors on the viability of their candidacy. Future empirical research might aim to document the extent of actors' anticipations about list performance and its incidence over aspects of candidate behaviour with respects to intra-party competition.

The second innovative aspect of the paper lies in the consideration of open- and flexible-list PR systems, as well as single- and multiple-preference forms of PLPR, within the same theoretical framework. As discussed, the prospect of broadening the scope of application of logical models beyond simple single-vote OLPR by taking into account multiple preference votes is something of a 'wager'. That is, we posited that – once we account for the different number of preferential votes used across systems – the intra-party allocation rules distinguishing OLPR from FLPR would not make a substantial difference. As the analysis presented in section 5.6.1 shows that the revised model works reasonably well across the FLPR and OLPR sub-samples (unlike the SBW model), we can tentatively conclude that the wager paid off. There is an important pragmatic rationale for extending the comparative analysis of intra-party competition to complex types of PLPR: these constitute the overwhelming majority of empirical cases. In Passarelli's (2020) review of PLPR systems, out of 29 countries for which this information is available, only seven employ single-preference OLPR (Finland, Kosovo, Poland, Brazil, Colombia, Indonesia, Lebanon), two of which have only recently switched from more complex systems (Indonesia from FLPR, and Kosovo from multiple-preference OLPR) and one of which allows parties to field closed lists if they wish (Colombia). In contrast, thirteen countries have $p > 1$ and fourteen use some form of FLPR for electing their national parliaments. In line with empirical research showing that prospects of advancement to electable positions produce a substantial degree of personal vote-seeking effort even under relatively restrictive FLPR rules (André et al., 2017), the findings of this paper should therefore encourage researchers to treat FLPR as a cognate of OLPR rather than 'closed lists in

disguise', at least as far as the aggregate outcomes of intra-party competition are concerned.

5.7.3 Limitations

All models are wrong, but some are useful. The reader hopefully will agree that the ones presented in this paper may belong to the latter category, insofar as they illustrate non-obvious relationships between electoral system quantities, have respectable predictive power, and – as discussed in this latter section – may have practical applications to institutional choice. Nonetheless, there are some areas of concern and related room for improvement; in particular, there are three outstanding issues with the present attempt to model intra-party quantities, which may be addressed in future research.

First, as noted in various parts of the paper, the models perform particularly poorly in the case of Slovakia, which employs a nationwide district of magnitude 150 (district magnitude M is equal to assembly size S). My sense is that this is due to a 'big fish' effect: politicians who are popular nationwide, like party leaders and prominent frontbenchers, can concentrate preference votes to an exceptional degree on their candidacy. This is of course happens beyond cases where $M = S$: the model over-predicts competitiveness just as poorly for the lists including incumbent Prime Ministers like Poland's Donald Tusk (75% of preference votes) or Belgium's Charles Michel (51%), as well as charismatic party leaders like Estonia's Martin Helme (94%) and Czechia's Tomio Okamura (48%). (And, given what a logical model is for, the model would have no business predicting these cases.) However, where $M \ll S$, these 'big fish' lists are only a fraction of the observations; whereas if $M = S$ then *all* lists will have at least a party leader on the ballot. The resulting bias of the prediction is thus substantially overstating competitiveness. Future iterations of a model for PLPR may thus attempt to account for the 'embeddedness' of districts in national politics: i.e. as the $\frac{M}{S}$ ratio tends to one, our prediction for v_1 should be adjusted upwards, and that for N_c should be adjusted downwards. At present, I am unable to justify a quantitative formalisation of such an adjustment.

Secondly, the theoretical derivation of v_s rests on the heroic assumption that mean and median of preference shares of seat eligible candidates are approximately equal. The empirical performance of the deriving models, both in terms of bias and precision, is respectable; but perhaps it is not respectable enough to warrant such a mathematical heresy, especially as the models for v_1 and N_c tend to do better than those for v_s, and to do so more consistently across specifications of the model and the sample. In sum, v_s remains the weakest link of the interlocked set of equations describing expected quantities of intra-party competition.

A final limitation of the argument presented is a clear definition of its scope conditions. As discussed,

there is value in extending the analysis to as broad a set of PLPR institutional set-ups as feasible, given the diversity of real-world variants of this electoral system family. In section 5.4, the reference to 'systems that are sufficiently proportional and sufficiently preferential' makes for an informal and intuitive way of defining the scope conditions of the theory, but I recognise that it is not sufficiently precise. A PLPR system with a very high representation threshold or a very small divisor formula might constrain inter-party competition enough to alter significantly actors' expectations over the number of seats at stake for an individual list. On the intra-party dimension, a FLPR system with an unattainably high preference threshold may make intra-party competition meaningless, and hence more random than the predictable patterns a logical model might aim for. Some of these factors may be accounted for by introducing further terms to the formulas, and I have made the case for sacrificing a degree of parsimony in favour of wider applicability to the complexity of real-world cases. But, for some extreme cases, we might simply have to conclude that we are dealing with something other than a *preferential* list *proportional* representation system.

5.7.4 Conclusion

In spite of these shortcomings, the empirical tests presented in the second part of the paper are overall rather encouraging for our quest towards a model of intra-party competition to match the seat-product model. Summing up, I have shown that we can predict from electoral and institutional quantities (1) what percentage of preference votes the top candidate will get, (2) how many candidates will effectively emerge within a list, and (3) what is the minimum share of preferences a candidate should get to be eligible for a seat. In the sample considered, these predictions are about as precise as those available for similar indicators of inter-party competition, they are less biased than those of existing intra-party models, and they perform consistently across sub-types of PLPR systems. Hopefully, as well as providing a stepping stone towards a more systematic and comprehensive research agenda on intra-party competition in PLPR, these results may also serve as testimony to the power of quantitative predictive logical models as a versatile and parsimonious theory-building tool.

Chapter 6

Conclusion

This thesis set out to advance and enrich the research agenda on the intra-party consequences of electoral institutions through the analysis of new, or relatively understudied, outcome variables. The three articles presented have shown how intra-party mechanisms link electoral systems to patterns of variation in democratic outcomes across countries, individual legislators, and party lists. At macro-level, the overall spatial representativeness of legislatures is affected by the way in which party nomination choices and voter leverage vary with constituency and ballot structure (Chapter 3). At micro-level, the career moves of individual legislators follow from the patterns of electoral security and accountability associated with district type (Chapter 4). At meso-level, it has been shown that the distribution of preference votes in PLPR systems is consistent with a theoretical framework that posits that district magnitude affects co-partisan competition through its relationships to both the number of co-partisans and to the expected number of seats attributed to a list (Chapter 5).

As well as studying political phenomena at three different levels of analysis, the three articles make contributions the study of electoral systems along different dimensions. The most valuable aspect of the first paper is, in my view, its *measurement* contribution: SURLI represents an original solution to the problem of comparing spatial descriptive representation across politically and geographically diverse polities. On the one hand, the methodological approach proposed be extended to other dimensions of spatial inequality, useful beyond the confines of electoral system studies. On the other hand, the measurement captures a dimension of descriptive representation – the representation of places – which is often pitted against other desired outcomes of electoral rules but which has so far proven elusive to gauge empirically. The second paper

makes an important *data* contribution, as it draws on data matching biographical, career, legislative and electoral information on a rare example of comparable sub-national legislatures elected via mixed-member systems: German State Parliaments. Beyond the specific case of level-hopping, the data could be further employed to shed light on the relationship between electoral institutions and other aspects of multi-level politics and legislative behaviour. The third article makes primarily a *theoretical* contribution to the study of electoral systems, extending the logical modelling approach to competition between co-partisans and presenting quantitative functional forms of the relationship between institutions and intra-party electoral outcomes. To the extent that electoral system research should aim for real-world relevance, the resulting predictive statements may help political actors to make informed institutional and strategic choices.

Scholars of electoral systems have traditionally concluded their works by stating perfunctorily that, while the literature has made progress on the inter-party dimension, not much else is known beyond how electoral institutions shape the allocation of seats to parties and the party system-level consequences of this relationship. There is of course still much work left to do, but such statement sounds increasingly like a cliché rather than a fair assessment of the literature. Contemporary electoral system scholarship has uncovered – and is uncovering – a great deal of subtle but consequential electoral system effects that operate through the way institutions shape incentives and opportunities for parties to manage their personnel, for politicians to advance their individual goals, and for voters to achieve forms of representation beyond party choice. Hopefully, the findings of this investigation have further added to these lines of enquiry. In my view, where we still fall short – this thesis included – is in translating this growing, cumulative knowledge into normative arguments. What weight should we give to local representation vis-á-vis other dimensions of descriptive and policy representation in designing electoral institutions? Should our electoral rules facilitate or prevent the vertical integration of political careers? How much, if any, competition between co-partisans is desirable? Perhaps, the future of electoral system research does not simply lie in 'finding out more' about the intra-party dimension (we are doing that). Rather, empirical enquiry should aim more explicitly at clarifying how intra-party consequences of electoral institutions should inform the normative trade-offs involved in institutional choice.

Bibliography

Aardal, Bernt. 2007. The Electoral System in Norway. In *The Evolution of Electoral and Party Systems in the Nordic Countries*, ed. Arend Lijphart and Bernard Grofman. Algora Publishing pp. 167–220.

Amorim Neto, Octavio and Gary W Cox. 1997. "Electoral institutions, cleavage structures, and the number of parties." *American Journal of Political Science* 41(1):149–174.

André, Audrey, André Freire and Zsófia Papp. 2014. Electoral rules and legislators' personal vote-seeking. In *Representing the People: A Survey Among Members of Statewide and Substate Parliaments*, ed. Kris Deschouwer and Sam Depauw. Oxford University Press pp. 87–109.

André, Audrey and Sam Depauw. 2013. "District magnitude and home styles of representation in European democracies." *West European Politics* 36(5):986–1006.

André, Audrey, Sam Depauw and Kris Deschouwer. 2014. "Legislators' local roots: Disentangling the effect of district magnitude." *Party Politics* 20(6):904–917.

André, Audrey, Sam Depauw and Matthew S Shugart. 2014. The Effect of Electoral Institutions on Legislative Behavior. In *Oxford Handbook of Legislative Studies*. Oxford University Press pp. 231–249.

André, Audrey, Sam Depauw, Matthew S Shugart and Roman Chytilek. 2017. "Party nomination strategies in flexible-list systems: Do preference votes matter?" *Party Politics* 23(5):589–600.

Appeldorn, Niels H and David Fortunato. 2021. "Legislative capacity in Germany's parliaments." *Legislative Studies Quarterly* 47(2):309–328.

Ariga, Kenichi. 2015. "Incumbency disadvantage under electoral rules with intraparty competition: evidence from Japan." *The Journal of Politics* 77(3):874–887.

Arter, David. 2013. "The 'Hows', not the 'Whys' or the 'Wherefores': The role of intra-party competition in the 2011 breakthrough of the True Finns." *Scandinavian Political Studies* 36(2):99–120.

Arzheimer, Kai and Jocelyn Evans. 2012. "Geolocation and voting: Candidate–voter distance effects on party choice in the 2010 UK general election in England." *Political Geography* 31(5):301–310.

Arzheimer, Kai and Jocelyn Evans. 2014. "Candidate geolocation and voter choice in the 2013 English County Council elections." *Research & Politics* 1(2):1–9.

Batto, Nathan F. 2012. "Differing mandates and party loyalty in mixed-member systems: Taiwan as a baseline case." *Electoral Studies* 31(2):384–392.

Baumann, Markus, Marc Debus and Tristan Klingelhöfer. 2017. "Keeping one's seat: the competitiveness of MP renomination in mixed-member electoral systems." *The Journal of Politics* 79(3):979–994.

Bernauer, Julian, Nathalie Giger and Jan Rosset. 2015. "Mind the gap: Do proportional electoral systems foster a more equal representation of women and men, poor and rich?" *International Political Science Review* 36(1):78–98.

Bieber, Ina and Luisa Wingerter. 2020. "Is It All a Question of the Electoral System? The Effects of Electoral System Types on the Representation of Women in German Municipal Councils." *German Politics* pp. 1–26.

Black, Gordon S. 1972. "A theory of political ambition: Career choices and the role of structural incentives." *American Political Science Review* 66(1):144–159.

Blais, André. 1991. "The debate over electoral systems." *International Political Science Review* 12(3):239–260.

Blais, André and Jean-François Daoust. 2017. "What do voters do when they like a local candidate from another party?" *Canadian Journal of Political Science/Revue canadienne de science politique* 50(4):1103–1109.

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

Blom-Hansen, Jens, Jørgen Elklit, Søren Serritzlew and Louise Riis Villadsen. 2016. "Ballot position and election results: Evidence from a natural experiment." *Electoral Studies* 44(1):172–183.

Borchert, Jens. 2011. "Individual ambition and institutional opportunity: A conceptual approach to political careers in multi-level systems." *Regional and Federal Studies* 21(2):117–140.

Borchert, Jens and Klaus Stolz. 2011*a*. "German political careers: The State level as an arena in its own right?" *Regional and Federal Studies* 21(2):205–222.

Borchert, Jens and Klaus Stolz. 2011*b*. "Institutional order and career patterns: Some comparative considerations." *Regional and Federal Studies* 21(2):271–282.

Borchert, Jens and Klaus Stolz. 2011*c*. "Introduction: Political careers in multi-level systems." *Regional and Federal Studies* 21(2):107–115.

Boucek, Francoise. 2009. "Rethinking factionalism: typologies, intra-party dynamics and three faces of factionalism." *Party politics* 15(4):455–485.

Bratton, Kathleen A. and Leonard P. Ray. 2002. "Descriptive representation, policy outcomes, and municipal day-care coverage in Norway." *American Journal of Political Science* 46(2):428–437.

Breunig, Christian, Emiliano Grossman and Miriam Hänni. 2022. "Responsiveness and democratic accountability: Observational evidence from an experiment in a mixed-member proportional system." *Legislative Studies Quarterly* 47(1):79–94.

British Columbia Citizens' Assembly on Electoral Reform. 2004. *Making Every Vote Count: The Case for Electoral Reform in British Columbia. Final Report.*

Brockington, David. 2004. "The paradox of proportional representation: The effect of party systems and coalitions on individuals' electoral participation." *Political Studies* 52(3):469–490.

Buisseret, Peter and Carlo Prato. 2018. "Legislative representation in flexible-list electoral systems." *Unpublished mansucript* .

Buisseret, Peter and Carlo Prato. 2020. "Voting behavior under proportional representation." *Journal of Theoretical Politics* 32(1):96–111.

Buisseret, Peter, Olle Folke, Carlo Prato and Johanna Rickne. 2022. "Party nomination strategies in list proportional representation systems." *American Journal of Political Science* 66(3):714–729.

Cain, Bruce, John Ferejohn and Morris Fiorina. 2013. *The Personal Vote.* Harvard University Press.

Calvo, Ernesto, Fernando Guarnieri and Fernando Limongi. 2015. "Why coalitions? Party system fragmentation, small party bias, and preferential vote in Brazil." *Electoral Studies* 39(1):219–229.

Campbell, Rosie and Philip Cowley. 2014. "What voters want: Reactions to candidate characteristics in a survey experiment." *Political Studies* 62(4):745–765.

Campbell, Rosie, Philip Cowley, Nick Vivyan and Markus Wagner. 2019. "Why friends and neighbors? Explaining the electoral appeal of local roots." *The Journal of Politics* 81(3):937–951.

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

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

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

Carey, John, Simon Hix, Mala Htun, Shaheen Mozaffar, G Bingham Powell and Andrew Reynolds. 2013. "Between science and engineering: Reflections on the APSA Presidential Task Force on political science, electoral rules, and democratic governance: Political scientists as electoral system engineers." *Perspectives on Politics* 11(3):827–840.

Carnes, Nicholas. 2012. "Does the numerical underrepresentation of the working class in congress matter?" *Legislative Studies Quarterly* 37(1):5–34.

Carnes, Nicholas and Noam Lupu. 2015. "Rethinking the comparative perspective on class and representation: Evidence from Latin America." *American Journal of Political Science* 59(1):1–18.

Carozzi, Felipe and Luca Repetto. 2016. "Sending the pork home: Birth town bias in transfers to Italian municipalities." *Journal of Public Economics* 134:42–52.

Catalinac, Amy. 2018. "Positioning under alternative electoral systems: Evidence from Japanese candidate election manifestos." *American Political Science Review* 112(1):31–48.

Centellas, Miguel. 2013. "Legislative turnover and institutional reforms: Evidence from the Bolivian case." *Paper prepared for presentation at the 2013 Annual Meeting of the American Political Science Association, Chicago, August 27–September 1, 2013.* .

Chang, Eric CC and Miriam A Golden. 2007. "Electoral systems, district magnitude and corruption." *British Journal of Political Science* 37(1):115–137.

Cheibub, José Antonio and Gisela Sin. 2020. "Preference vote and intra-party competition in open list PR systems." *Journal of Theoretical Politics* 32(1):70–95.

Cheibub, José Antonio and Monika Nalepa. 2020. "Revisiting electoral personalism." *Journal of Theoretical Politics* 32(1):3–10.

Childs, Sarah and Philip Cowley. 2011. "The politics of local presence: Is there a case for descriptive representation?" *Political Studies* 59(1):1–19.

Cohen, Scott and Leonidas Guibas. 1997. The earth mover's distance: Lower bounds and invariance under translation. Technical report Stanford University (CA) Department of Computer Science.

Coman, Emanuel Emil. 2012. "Legislative behavior in Romania: The effect of the 2008 Romanian electoral reform." *Legislative Studies Quarterly* 37(2):199–224.

Cowley, Philip. 2013. "Why not ask the audience? Understanding the public's representational priorities." *British Politics* 8(2):138–163.

Cox, Gary W. 1996. "Is the single nontransferable vote superproportional? Evidence from Japan and Taiwan." *American Journal of Political Science* 40(3):740–755.

Cox, Gary W. 1997. *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge University Press.

Cox, Gary W. 2015. "Electoral rules, mobilization, and turnout." *Annual Review of Political Science* 18(1):49–68.

Cox, Gary W, Jon H Fiva and Daniel M Smith. 2016. "The contraction effect: How proportional representation affects mobilization and turnout." *The Journal of Politics* 78(4):1249–1263.

Cox, Karen E and Leonard J Schoppa. 2002. "Interaction effects in mixed-member electoral systems: theory and evidence from Germany, Japan, and Italy." *Comparative Political Studies* 35(9):1027–1053.

Crisp, Brian F, Kathryn M Jensen and Yael Shomer. 2007. "Magnitude and vote seeking." *Electoral Studies* 26(4):727–734.

Crisp, Brian F and William M Simoneau. 2018. Electoral systems and constituency service. In *The Oxford Handbook of Electoral Systems*, ed. Erik S Herron, Robert S. Pekkanen and Matthew S. Shugart. Oxford University Press pp. 345–363.

Curtice, John and Phillips Shively. 2009. Who Represents Us Best? One Member or Many? In *The Comparative Study of Electoral Systems*, ed. Hans-Dieter Klingemann. Oxford University Press pp. 171–192.

Cuturi, Marco. 2013. "Sinkhorn distances: Lightspeed computation of optimal transport." *Advances in neural information processing systems* 26:2292–2300.

Dancygier, Rafaela M. 2017. *Dilemmas of Inclusion*. Princeton University Press.

Dassonneville, Ruth and Ian McAllister. 2018. "Gender, political knowledge, and descriptive representation: The impact of long-term socialization." *American Journal of Political Science* 62(2):249–265.

Däubler, Thomas. 2020. "Do more flexible lists increase the take-up of preference voting?" *Electoral Studies* 68:102232.

Däubler, Thomas and Simon Hix. 2018. "Ballot structure, list flexibility and policy representation." *Journal of European Public Policy* 25(12):1798–1816.

De Montesquieu, Charles. 1989. *Montesquieu: The Spirit of the Laws*. Cambridge University Press.

Detterbeck, Klaus. 2012. *Multi-level Party Politics in Western Europe*. Vol. 2 Springer.

Detterbeck, Klaus. 2016. "Candidate selection in Germany: Local and regional party elites still in control?" *American Behavioral Scientist* 60(7):837–852.

Devroe, Robin and Bram Wauters. 2020. "Does high on the ballot means highly competent? Explaining the ballot position effect in list-PR systems." *Acta Politica* 55(3):454–471.

Dhima, Kostanca, Sona N Golder, Laura B Stephenson and Karine Van der Straeten. 2021. "Permissive electoral systems and descriptive representation." *Electoral Studies* 73:102381.

Di Capua, Roberto, Andrea Pilotti, André Mach and Karim Lasseb. 2022. "Political professionalization and transformations of political career patterns in multi-level states: The case of Switzerland." *Regional & Federal Studies* 32(1):95–114.

Docherty, David. 2011. "The Canadian political career structure: From stability to free agency." *Regional and Federal Studies* 21(2):185–203.

Dodeigne, Jérémy. 2014. "(Re-) Assessing career patterns in multi-level systems: Insights from Wallonia in Belgium." *Regional & Federal Studies* 24(2):151–171.

Dodeigne, Jérémy and Jean-Benoit Pilet. 2021. "Centralized or decentralized personalization? Measuring intra-party competition in open and flexible list PR systems." *Party Politics* 27(2):234–245.

Droop, Henry Richmond. 1881. "On methods of electing representatives." *Journal of the Statistical Society of London* 44(2):141–202.

Dudzińska, Agnieszka, Corentin Poyet, Olivier Costa and Bernhard Weßels. 2014. Representational Roles. In *Representing the People: A Survey among Members of Statewide and Sub-state Parliaments*. Oxford University Press pp. 19–38.

Duflo, Esther. 2012. "Women empowerment and economic development." *Journal of Economic literature* 50(4):1051–79.

Duverger, Maurice. 1959. *Political parties: Their Organization and Activity in the Modern State*. Metheun & Co. Ltd.

Dynes, Adam M, Hans JG Hassell and Matthew R Miles. 2019. "The personality of the politically ambitious." *Political Behavior* 41(2):309–336.

Edinger, Michael and Stefan Jahr. 2016. *Political Careers in Europe: Career Patterns in Multi-level Systems*. Bloomsbury Publishing.

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

Evans, Jocelyn, Kai Arzheimer, Rosie Campbell and Philip Cowley. 2017. "Candidate localness and voter choice in the 2015 General Election in England." *Political Geography* 59:61–71.

Farrell, David M and Ian McAllister. 2006. "Voter satisfaction and electoral systems: Does preferential voting in candidate-centred systems make a difference?" *European Journal of Political Research* 45(5):723–749.

Ferrara, Federico. 2004. "Electoral coordination and the strategic desertion of strong parties in compensatory mixed systems with negative vote transfers." *Electoral Studies* 23(3):391–413.

Ferrara, Federico, Erik Herron and Misa Nishikawa. 2005. *Mixed Electoral Systems: Contamination and its Consequences*. Springer.

Ferrara, Federico and Erik S Herron. 2005. "Going it alone? Strategic entry under mixed electoral rules." *American Journal of Political Science* 49(1):16–31.

Fiva, Jon H, Askill Harkjerr Halse and Daniel M Smith. 2018. "Local Candidates and Distributive Politics under Closed-List Proportional Representation." *CESifo Working Paper No. 7039* .

Francis, Wayne L and Lawrence W Kenny. 1996. "Position shifting in pursuit of higher office." *American Journal of Political Science* 40(3):768–786.

Freedom House. 2012. *Freedom in the World 2012: The Annual Survey of Political Rights and Civil Liberties*. Rowman & Littlefield.

Friedman, Eben. 2005. "Electoral system design and minority representation in Slovakia and Macedonia." *Ethnopolitics* 4(4):381–396.

Fulton, Sarah A, Cherie D Maestas, L Sandy Maisel and Walter J Stone. 2006. "The sense of a woman: Gender, ambition, and the decision to run for congress." *Political Research Quarterly* 59(2):235–248.

Gallagher, Michael. 1980. "Candidate selection in Ireland: The impact of localism and the electoral system." *British Journal of Political Science* 10(4):489–503.

Garand, James C. 1988. "Localism and regionalism in presidential elections: Is there a home state or regional advantage?" *Western Political Quarterly* 41(1):85–103.

Geese, Lucas and Diana Schacht. 2019. "The more concentrated, the better represented? The geographical concentration of immigrants and their descriptive representation in the German mixed-member system." *International Political Science Review* 40(5):643–658.

Gerring, John, Erzen Oncel, Kevin Morrison and Philip Keefer. 2014. "The global leadership project: A comprehensive database of political elites." *Available at SSRN 2491672* .

Gimpel, James G, Kimberly A Karnes, John McTague and Shanna Pearson-Merkowitz. 2008. "Distance-decay in the political geography of friends-and-neighbors voting." *Political Geography* 27(2):231–252.

Göbel, Sascha and Simon Munzert. 2022. "The comparative legislators database." *British Journal of Political Science* 52(3):1398–1408.

Golder, Matt and Jacek Stramski. 2010. "Ideological congruence and electoral institutions." *American Journal of Political Science* 54(1):90–106.

Gonzalez-Eiras, Martin and Carlos Sanz. 2021. "Women's representation in politics: The effect of electoral systems." *Journal of Public Economics* 198:104399.

Górecki, Maciej A and Michael Marsh. 2012. "Not just 'friends and neighbours': Canvassing, geographic proximity and voter choice." *European Journal of Political Research* 51(5):563–582.

Grofman, Bernard. 2005. "Comparisons among electoral systems: distinguishing between localism and candidate-centered politics." *Electoral Studies* 24(4):735–740.

Grofman, Bernard, Sung-Chull Lee, Edwin Winckler and Brian Woodall. 1999. *Elections in Japan, Korea, and Taiwan under the Single Non-transferable Vote: The Comparative Study of an Embedded Institution.* University of Michigan Press.

Gschwend, Thomas, Ron Johnston and Charles Pattie. 2003. "Split-ticket patterns in mixed-member proportional election systems: Estimates and analyses of their spatial variation at the German federal election, 1998." *British Journal of Political Science* 33(1):109–127.

Hainmueller, Jens and Holger Lutz Kern. 2008. "Incumbency as a source of spillover effects in mixed electoral systems: Evidence from a regression-discontinuity design." *Electoral Studies* 27(2):213–227.

Hallerberg, Mark. 2004. Electoral laws, government, and parliament. In *Patterns of Parliamentary Behaviour. Passage of Legislation across Western Europe*, ed. Herbert Döring and Mark Hallerberg. Routledge pp. 11–33.

Hamilton, Alexander, James Madison and John Jay. 2008. *The Federalist Papers.* Oxford University Press.

Heath, Oliver. 2018. "Policy alienation, social alienation and working-class abstention in Britain, 1964–2010." *British Journal of Political Science* 48(4):1053–1073.

Heinsohn, Till. 2014. "Institutional determinants of legislative turnover in the German state parliaments: 1947–2012." *The Journal of Legislative Studies* 20(4):473–494.

Heinsohn, Till and Melissa Schiefer. 2019. "Advancing to positions of power in parliament – does seniority matter?" *The Journal of Legislative Studies* 25(4):511–532.

Herrick, Rebekah and Michael K Moore. 1993. "Political ambition's effect on legislative behavior: Schlesinger's typology reconsidered and revisited." *The Journal of Politics* 55(3):765–776.

Herron, Erik S. 2002. "Electoral influences on legislative behavior in mixed-member systems: Evidence from Ukraine's Verkhovna Rada." *Legislative Studies Quarterly* 27(3):361–382.

Herron, Erik S, Kuniaki Nemoto and Misa Nishikawa. 2018. Reconciling approaches in the study of mixed-member electoral systems. In *The Oxford Handbook of Electoral Systems*, ed. Erik S Herron, Robert S. Pekkanen and Matthew S. Shugart. Oxford University Press pp. 445–472.

Herron, Erik S and Misa Nishikawa. 2001. "Contamination effects and the number of parties in mixed-superposition electoral systems." *Electoral Studies* 20(1):63–86.

Hix, Simon. 2004. "Electoral institutions and legislative behavior: Explaining voting defection in the European Parliament." *World Politics* 56(2):194–223.

Horiuchi, Yusaku, Daniel M Smith and Teppei Yamamoto. 2018. "Identifying voter preferences for politicians' personal attributes: A conjoint experiment in Japan." *Political Science Research and Methods* 8(1):75–91.

Høyland, Bjørn, Sara B Hobolt and Simon Hix. 2019. "Career ambitions and legislative participation: The moderating effect of electoral institutions." *British Journal of Political Science* 49(2):491–512.

Htun, Mala. 2005. Women, political parties and electoral systems in Latin America. In *Women in parliament: Beyond numbers*, ed. Julie Ballington and Azza Karam. International IDEA Stockholm, Sweden pp. 112–121.

Invernizzi, Giovanna M. 2021. "Antagonistic cooperation: Factional competition in the shadow of elections." *American Journal of Political Science,* Early View.

Itzkovitch-Malka, Reut and Reuven Y Hazan. 2017. "Unpacking party unity: The combined effects of electoral systems and candidate selection methods on legislative attitudes and behavioural norms." *Political Studies* 65(2):452–474.

Iversen, Torben and David Soskice. 2006. "Electoral institutions and the politics of coalitions: Why some democracies redistribute more than others." *American Political Science Review* 100(2):165–181.

Jahr, Stefan. 2015. Career movements across parliamentary levels: the German case. In *Political Careers in Europe. Career Patterns in Multi-Level Systems*, ed. Michael Edinger and Stefan Jahr. Bloomsbury Publishing pp. 55–75.

Janssen, Chloé. 2022. "Shaping the (dis) advantage: the impact of partisan and demographic factors on ethnic minority candidates' success in preferential voting systems. Evidence from the Brussels case." *Journal of Elections, Public Opinion and Parties* 32(1):22–45.

Jenkins, Lord Roy. 1998. *The Report of the Independent Commission on the Voting System.* Stationery Office.

Jennes, Geert and Damiaan Persyn. 2015. "The effect of political representation on the geographic distribution of income: Evidence using Belgian data." *European Journal of Political Economy* 37(1):178–194.

Johnson, Catherine and Gemma Rosenblatt. 2006. "Do MPs have the 'right stuff'?" *Parliamentary Affairs* 60(1):164–169.

Jones, Mark P, Sebastian Saiegh, Pablo T Spiller and Mariano Tommasi. 2002. "Amateur legislators–professional politicians: The consequences of party-centered electoral rules in a federal system." *American Journal of Political Science* 46(3):656–669.

Joshi, Devin K. 2013. "The representation of younger age cohorts in Asian parliaments: Do electoral systems make a difference?" *Representation* 49(1):1–16.

Jun, Hae-Won and Simon Hix. 2010. "Electoral systems, political career paths and legislative behavior: Evidence from South Korea's mixed-member system." *Japanese Journal of Political Science* 11(2):153–171.

Karvonen, Lauri. 2004. "Preferential voting: Incidence and effects." *International Political Science Review* 25(2):203–226.

Kelly, Richard. 2008. "It's only made things worse: a critique of electoral reform in Britain." *The Political Quarterly* 79(2):260–268.

Kerevel, Yann. 2010. "The legislative consequences of Mexico's mixed-member electoral system, 2000–2009." *Electoral Studies* 29(4):691–703.

Key, Valdimer O. 1949. *Southern Politics*. Random House.

Kim, Chong Lim. 1970. "Political attitudes of defeated candidates in an American state election." *American Political Science Review* 64(3):879–887.

Kimura, Daniel K. 1992. "Symmetry and scale dependence in functional relationship regression." *Systematic Biology* 41(2):233–241.

Kostadinova, Tatiana. 2007. "Ethnic and women's representation under mixed election systems." *Electoral Studies* 26(2):418–431.

Kselman, Daniel M. 2020. "Public goods equilibria under closed-and open-list proportional representation." *Journal of Theoretical Politics* 32(1):112–142.

Laakso, Markku and Rein Taagepera. 1979. ""Effective" number of parties: a measure with application to West Europe." *Comparative Political Studies* 12(1):3–27.

Ladner, Andreas, Nicolas Keuffer and Harald Baldersheim. 2016. "Measuring local autonomy in 39 countries (1990–2014)." *Regional & Federal Studies* 26(3):321–357.

Lancaster, Thomas D and W David Patterson. 1990. "Comparative pork barrel politics: perceptions from the West German Bundestag." *Comparative Political Studies* 22(4):458–477.

Lazarus, Jeffrey. 2006. "Term limits' multiple effects on state legislators' career decisions." *State Politics & Policy Quarterly* 6(4):357–383.

Lessmann, Christian and André Seidel. 2017. "Regional inequality, convergence, and its determinants–A view from outer space." *European Economic Review* 92(1):110–132.

Li, Yuhui and Matthew S Shugart. 2016. "The seat product model of the effective number of parties: A case for applied political science." *Electoral Studies* 41(1):23–34.

Lijphart, Arend. 1984. *Democracies: Patterns of Majoritarian and Consensus Government in Twenty-one Countries*. Yale University Press.

Lijphart, Arend. 1985. "The field of electoral systems research: A critical survey." *Electoral Studies* 4(1):3–14.

Lijphart, Arend. 2012. *Patterns of Democracy*. Yale University Press.

Lo Russo, Michele and Luca Verzichelli. 2016. Reshaping political careers in post-transition Italy: A synchronic analysis. In *Political Careers in Europe: Career Patterns in Multi-level Systems*, ed. Michael Edinger and Stefan Jahr. Bloomsbury Publishing pp. 27–54.

Lovenduski, Joni and Pippa Norris. 2003. "Westminster women: the politics of presence." *Political Studies* 51(1):84–102.

Lublin, David. 2017. "Electoral systems, ethnic heterogeneity and party system fragmentation." *British Journal of Political Science* 47(2):373–389.

Lundberg, Thomas Carl. 2006. "Second-class representatives? Mixed-member proportional representation in Britain." *Parliamentary Affairs* 59(1):60–77.

Lupu, Noam, Lucía Selios and Zach Warner. 2017. "A new measure of congruence: The Earth Mover's Distance." *Political Analysis* 25(1):95–113.

Lutz, Georg. 2010. "First come, first served: The effect of ballot position on electoral success in open ballot PR elections." *Representation* 46(2):167–181.

Maaser, Nicola and Thomas Stratmann. 2018. "Election rules, legislators' incentives, and policy outcomes: Evidence from the mixed member system in Germany." *European Journal of Political Economy* 54(1):227–239.

Maestas, Cherie. 2003. "The incentive to listen: Progressive ambition, resources, and opinion monitoring among state legislators." *Journal of Politics* 65(2):439–456.

Manow, Philip. 2007. "Electoral rules and legislative turnover: Evidence from Germany's mixed electoral system." *West European Politics* 30(1):195–207.

Mansbridge, Jane. 1999. "Should blacks represent blacks and women represent women? A contingent "yes"." *The Journal of Politics* 61(3):628–657.

Marsh, Michael. 1985. "The voters decide?: Preferential voting in European list systems." *European Journal of Political Research* 13(4):365–378.

Marsh, Michael. 1988. *Candidate selection in comparative perspective: The secret garden of politics*. SAGE Publications Limited.

Marsh, Michael. 2007. "Candidates or parties? Objects of electoral choice in Ireland." *Party Politics* 13(4):500–527.

Martin, Shane. 2014. "Why electoral systems don't always matter: The impact of 'mega-seats' on legislative behaviour in Ireland." *Party Politics* 20(3):467–479.

Massicotte, Louis. 2003. "To create or to copy? Electoral systems in the German Länder." *German Politics* 12(1):1–22.

Massicotte, Louis, André Blais and Antoine Yoshinaka. 2004. *Establishing the Rules of the Game: Election Laws in Democracies*. University of Toronto Press.

Matland, Richard E and Donley T Studlar. 1996. "The contagion of women candidates in single-member district and proportional representation electoral systems: Canada and Norway." *The Journal of Politics* 58(3):707–733.

Matland, Richard E and Donley T Studlar. 2004. "Determinants of legislative turnover: A cross-national analysis." *British Journal of Political Science* 34(1):87–108.

Mayhew, David R. 1974. "Congressional elections: The case of the vanishing marginals." *Polity* 6(3):295–317.

McAllister, Ian and Donley T Studlar. 2002. "Electoral systems and women's representation: a long-term perspective." *Representation* 39(1):3–14.

Milazzo, Caitlin and Joshua Townsley. 2018. "Conceived in Harlesden: Candidate-centred campaigning in British general elections." *Parliamentary Affairs* 73(1).

Minicucci, Stephen. 2001. "The "cement of interest": Interest-based models of nation-building in the early republic." *Social Science History* 25(2):247–274.

Morgenstern, Scott. 2003. *Patterns of legislative politics: Roll-call voting in Latin America and the United States*. Cambridge University Press.

Morgenstern, Scott and Stephen M Swindle. 2005. "Are politics local? An analysis of voting patterns in 23 democracies." *Comparative Political Studies* 38(2):143–170.

Moser, Robert G. 2008. "Electoral systems and the representation of ethnic minorities: Evidence from Russia." *Comparative Politics* 40(3):273–292.

Moser, Robert G and Ethan Scheiner. 2012. *Electoral Systems and Political Context: How the Effects of Rules Vary Across New and Established Democracies*. Cambridge University Press.

Nemčok, Miroslav and Jakub Šedo. 2018. "Theoretical expectations and actual outcomes of electoral systems: how to measure the size of the deviation?" *World Political Science* 14(2):189–212.

Norris, Pippa. 2004. *Electoral Engineering: Voting Rules and Political Behavior*. Cambridge university press.

Norris, Pippa. 2006. "The impact of electoral reform on women's representation." *Acta Politica* 41(2):197–213.

Ohmura, Tamaki. 2014. "When your name is on the list, it is time to party: the candidacy divide in a mixed-member proportional system." *Representation* 50(1):69–82.

Olivella, Santiago and Margit Tavits. 2014. "Legislative effects of electoral mandates." *British Journal of Political Science* 44(2):301–321.

Oñate, Pablo. 2018. Political careers in Spain: Mobility between political arenas in a multilevel system. In *Borders and Margins: Federalism, Devolution and Multi-Level Governance*, ed. Guy Lachapelle and Pablo Oñate. Barbara Budrich Publishers pp. 245–258.

Pachón, Mónica and Matthew S Shugart. 2010. "Electoral reform and the mirror image of inter-party and intra-party competition: The adoption of party lists in Colombia." *Electoral Studies* 29(4):648–660.

Panagopoulos, Costas, Jan E Leighley and Brian T Hamel. 2017. "Are voters mobilized by a 'friend-and-neighbor' on the ballot? Evidence from a field experiment." *Political Behavior* 39(4):865–882.

Passarelli, Gianluca. 2020. *Preferential Voting Systems*. Springer.

Patzelt, Werner J. 2007. "The constituency roles of MPs at the federal and Länder levels in Germany." *Regional and Federal Studies* 17(1):47–70.

Pedersen, Mogens N, Ulrik Kjaer and Kjell Eliassen. 2007. The geographical dimension of parliamentary recruitment: among native sons and parachutists. In *Democratic Representation in Europe: Diversity, Change, and Convergence*, ed. Maurizio Cotta and Heinrich Best. Oxford University Press pp. 160–190.

Persson, Torsten and Guido Tabellini. 2005. *The Economic Effects of Constitutions*. MIT press.

Pitkin, Hanna Fenichel. 1967. *The Concept of Representation*. University of California Press.

Portmann, Lea and Nenad Stojanović. 2019. "Electoral discrimination against immigrant-origin candidates." *Political Behavior* 41(1):105–134.

Powell Jr, G Bingham and Georg S Vanberg. 2000. "Election laws, disproportionality and median correspondence: Implications for two visions of democracy." *British Journal of Political Science* 30(3):383–411.

Preuhs, Robert R. 2005. "Descriptive representation, legislative leadership, and direct democracy: Latino influence on English only laws in the States, 1984–2002." *State Politics & Policy Quarterly* 5(3):203–224.

Rae, Douglas W. 1967. *The Political Consequences of Electoral Laws*. Yale University Press.

Ragauskas, Rimvydas and Frank Thames. 2020. "Cross-tier personal gains in mixed electoral systems." *Journal of Elections, Public Opinion and Parties* pp. 1–23.

Reeve, Andrew and Alan Ware. 2013. *Electoral Systems: A Theoretical and Comparative Introduction*. Routledge.

Rehfeld, Andrew. 2005. *The Concept of Constituency: Political Eepresentation, Democratic Legitimacy, and Institutional Design*. Cambridge University Press.

Reiser, Marion. 2014. "The universe of group representation in Germany: Analysing formal and informal party rules and quotas in the process of candidate selection." *International Political Science Review* 35(1):55–66.

Renwick, Alan and Jean-Benoit Pilet. 2016. *Faces on the Ballot: The Personalization of Electoral Systems in Europe*. Oxford University Press.

Reutter, Werner. 2006. "The transfer of power hypothesis and the German Länder: In need of modification." *Publius: The Journal of Federalism* 36(2):277–301.

Reutter, Werner. 2021. *The German Länder: An Introduction*. Springer.

Rice, Tom W and Alisa A Macht. 1987. "The hometown advantage: mobilization or conversion?" *Political Behavior* 9(3):257–262.

Rich, Timothy S. 2014. "Party Voting Cohesion in Mixed Member Legislative Systems: Evidence from Korea and Taiwan." *Legislative Studies Quarterly* 39(1):113–135.

Riera, Pedro and Francisco Cantú. 2022. "Electoral systems and ideological voting." *European Political Science Review* pp. 1–19.

Robbins, Joseph W. 2010. "The personal vote and voter turnout." *Electoral Studies* 29(4):661–672.

Roberts, Geoffrey. 1988. German Federal Republic: Two lane route to Bonn. In *Candidate Selection in Comparative Perspective: The Secret Garden of Politics. London*, ed. Michael Marsh. SAGE Publications Ltd pp. 94–118.

Roeder, Philip G. 2009. "Ethnofederalism and the mismanagement of conflicting nationalisms." *Regional & Federal Studies* 19(2):203–219.

Rohde, David W. 1979. "Risk-bearing and progressive ambition: The case of members of the United States House of Representatives." *American Journal of Political Science* 32(1):1–26.

Roy, Jason and Christopher Alcantara. 2015. "The candidate effect: Does the local candidate matter?" *Journal of Elections, Public Opinion & Parties* 25(2):195–214.

Rubner, Yossi, Carlo Tomasi and Leonidas J Guibas. 2000. "The Earth Mover's Distance as a metric for image retrieval." *International Journal of Computer Vision* 40(2):99–121.

Ruchelman, Leonard I. 1970. *Political Careers: Recruitment Through the Legislature*. Fairleigh Dickinson University Press.

Rudolph, Lukas and Thomas Däubler. 2016. "Holding individual representatives accountable: the role of electoral systems." *The Journal of Politics* 78(3):746–762.

Rule, Wilma. 1994. Parliaments of, by, and for the people: Except for women? In *Electoral Systems in Comparative Perspective: Their Impact on Women and Minorities*, ed. Joseph F. Zimmerman. Greenwood Press pp. 15–30.

Sällberg, Yohanna and Martin Ejnar Hansen. 2020. "Analysing the Importance of Localness for MP Campaigning and Legislative Performance." *Representation* 56(2):1–13.

Samuels, David. 2003. *Ambition, Federalism, and Legislative Politics in Brazil*. Cambridge University Press.

Santos, Fabiano GM and Fabiano JH Pegurier. 2011. "Political careers in Brazil: long-term trends and cross-sectional variation." *Regional and Federal Studies* 21(2):165–183.

Sartori, Giovanni. 1968. "Political development and political engineering." *Public Policy* 17(1):261–298.

Schlesinger, Joseph. 1966. *Ambition and Politics: Political Careers in the United States*. Rand McNally and Co.

Schmidt, Gregory D. 2009. "The election of women in list PR systems: Testing the conventional wisdom." *Electoral Studies* 28(2):190–203.

Schwindt-Bayer, Leslie A. 2005. "The incumbency disadvantage and women's election to legislative office." *Electoral Studies* 24(2):227–244.

Selb, Peter and Georg Lutz. 2015. "Lone fighters: Intraparty competition, interparty competition, and candidates' vote seeking efforts in open-ballot PR elections." *Electoral Studies* 39(1):329–337.

Shomer, Yael. 2017. "The conditional effect of electoral systems and intraparty candidate selection processes on parties' behavior." *Legislative Studies Quarterly* 42(1):63–96.

Shugart, Matthew and Martin P Wattenberg. 2001. *Mixed-member Electoral Systems: The Best of Both Worlds?* OUP Oxford.

Shugart, Matthew S. 2005. Comparative electoral systems research: The maturation of a field and new challenges ahead. In *The Politics of Electoral Systems*, ed. Michael Gallagher and Paul Mitchell. Oxford University Press pp. 25–56.

Shugart, Matthew S, Matthew E Bergman and Kevin A Watt. 2013. "Patterns of intraparty competition in open-list & SNTV systems." *Electoral Studies* 32(2):321–333.

Shugart, Matthew S and Rein Taagepera. 1989. *Seats and Votes: The Effects & Determinants of Electoral Systems*. Yale University Press.

Shugart, Matthew S and Rein Taagepera. 2017. *Votes from Seats: Logical Models of Electoral Systems*. Cambridge University Press.

Shugart, Matthew Søberg, Melody Ellis Valdini and Kati Suominen. 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.

Sikk, Allan and Rein Taagepera. 2014. "How population size affects party systems and cabinet duration." *Party Politics* 20(4):591–603.

Söderlund, Peter. 2017. "Candidate-centred electoral systems and voter turnout." *West European Politics* 40(3):516–533.

Squire, Peverill. 1988. "Career opportunities and membership stability in legislatures." *Legislative Studies Quarterly* pp. 65–82.

Squire, Peverill. 1992. "Legislative professionalization and membership diversity in state legislatures." *Legislative Studies Quarterly* 17(1):69–79.

Stoffel, Michael F. 2014. "MP behavior in mixed-member electoral systems." *Electoral Studies* 35:78–87.

Stolz, Klaus. 2011. "The regionalization of political careers in Spain and the UK." *Regional and Federal Studies* 21(2):223–243.

Stolz, Klaus. 2013. *Towards a Regional Political Class?: Professional Politicians and Regional Institutions in Catalonia and Scotland*. Manchester University Press.

Stolz, Klaus. 2015. Legislative Careers in a Multi-Level Europe. In *Political Careers in Europe*, ed. Michael Edinger and Stefan Jahr. Nomos Verlagsgesellschaft mbH & Co. KG pp. 179–204.

Stratmann, Thomas and Martin Baur. 2002. "Plurality rule, proportional representation, and the German Bundestag: How incentives to pork-barrel differ across electoral systems." *American Journal of Political Science* 46(3):506–514.

Swenden, Wilfried. 2006. *Federalism and Regionalism in Western Europe: A Comparative and Thematic Analysis*. Springer.

Taagepera, Rein. 2007. *Predicting party sizes: The logic of simple electoral systems*. OUP Oxford.

Taagepera, Rein. 2008. *Making social sciences more scientific: The need for predictive models*. Oxford University Press.

Taagepera, Rein and Allan Sikk. 2010. "Parsimonious model for predicting mean cabinet duration on the basis of electoral system." *Party Politics* 16(2):261–281.

Taagepera, Rein and Bernard Grofman. 1985. "Rethinking Duverger's law: predicting the effective number of parties in plurality and PR systems–parties minus issues equals one." *European Journal of Political Research* 13(4):341–352.

Taagepera, Rein and Matthew S Shugart. 1993. "Predicting the number of parties: A quantitative model of Duverger's mechanical effect." *American Political Science Review* 87(2):455–464.

Taagepera, Rein, Peter Selb and Bernard Grofman. 2014. "How turnout depends on the number of parties: A logical model." *Journal of Elections, Public Opinion & Parties* 24(4):393–413.

Tavits, Margit. 2010. "Effect of local ties on electoral success and parliamentary behaviour: The case of Estonia." *Party Politics* 16(2):215–235.

Thames, Frank C. 2005. "A house divided: party strength and the mandate divide in Hungary, Russia, and Ukraine." *Comparative Political Studies* 38(3):282–303.

Valdini, Melody Ellis. 2012. "A deterrent to diversity: The conditional effect of electoral rules on the nomination of women candidates." *Electoral Studies* 31(4):740–749.

Van Erkel, Patrick FA and Peter Thijssen. 2016. "The first one wins: Distilling the primacy effect." *Electoral Studies* 44(1):245–254.

Van Erkel, Patrick FA, Peter Thijssen and Peter Van Aelst. 2017. "One for all or all for one: The electoral effects of personalized campaign strategies." *Acta Politica* 52(3):384–405.

Vanlangenakker, Ine, Bart Maddens and Gert-Jan Put. 2013. "Career patterns in multilevel states: An analysis of the Belgian regions." *Regional Studies* 47(3):356–367.

Vowles, Jack. 2015. "Legislative accountability in a mixed-member system." *Australian Journal of Political Science* 50(2):279–296.

Wagner, Wolfgang. 2014. "The overstated merits of proportional representation: The Republic of Macedonia as a natural experiment for assessing the impact of electoral systems on descriptive representation." *Ethnopolitics* 13(5):483–500.

Wängnerud, Lena. 2009. "Women in parliaments: Descriptive and substantive representation." *Annual Review of Political Science* 12:51–69.

Weber, Max. 2008. *Max Weber's Complete Writings on Academic and Political Vocations*. Algora Publishing.

Wessels, Bernhard. 1999. Whom to Represent? Role Orientations of Legislators in Europe. In *Political Representation and Legitimacy in the European Union*, ed. Hermann Schmitt and Jacques Thomassen. Oxford University Press.

World Bank, Database. 2010. *World Development Indicators*. Washington, DC: World Bank.

Zittel, Thomas and Thomas Gschwend. 2008. "Individualised constituency campaigns in mixed-member electoral systems: Candidates in the 2005 German elections." *West European Politics* 31(5):978–1003.

Zollinger, Daniel and Daniel Bochsler. 2012. "Minority representation in a semi-democratic regime: the Georgian case." *Democratization* 19(4):611–641.

Appendix

A First Paper: Validation of EMD Proxy

This section illustrates our approach to deriving a proxy for two-dimensional Earth Mover's Distance (EMD) measure of discrepancy between two spatial distributions. We begin with two distributions, each characterized by a set of coordinates in two dimensions and associated weights, with each distribution's weights summing to 1. In one dimension, the EMD is equivalent to the integral of the discrepancy between the two cumulative distribution functions (CDFs), and can thus be computed quickly. In more than one dimension, the EMD is computationally costly and thus inconvenient for distributions with many coordinates.

Our proposed proxy computes the EMD in one dimension, then repeats the calculation over several rotations of the data, and finally averages these measurements. Figure A.1 below conveys the concept: we sweep through the data in the direction of each arrow, computing the 1-dimensional EMD (equivalently, the integral of CDF discrepancy) in each pass – in the figure, the cases for 3 and 6 rotations are shown – and then average the values of the EMD obtained in each of these passes.

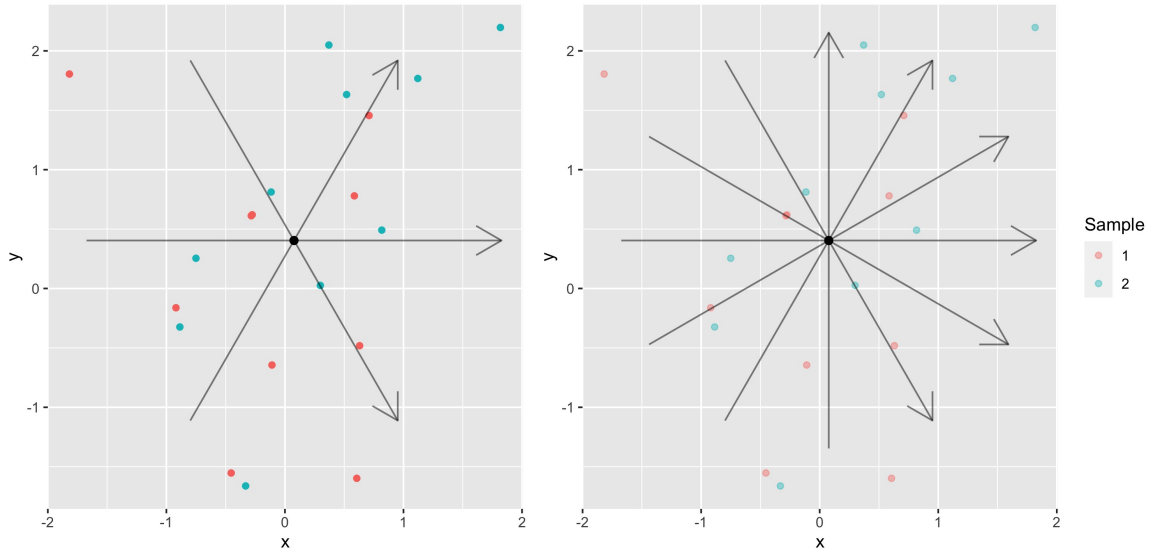


Figure A.1

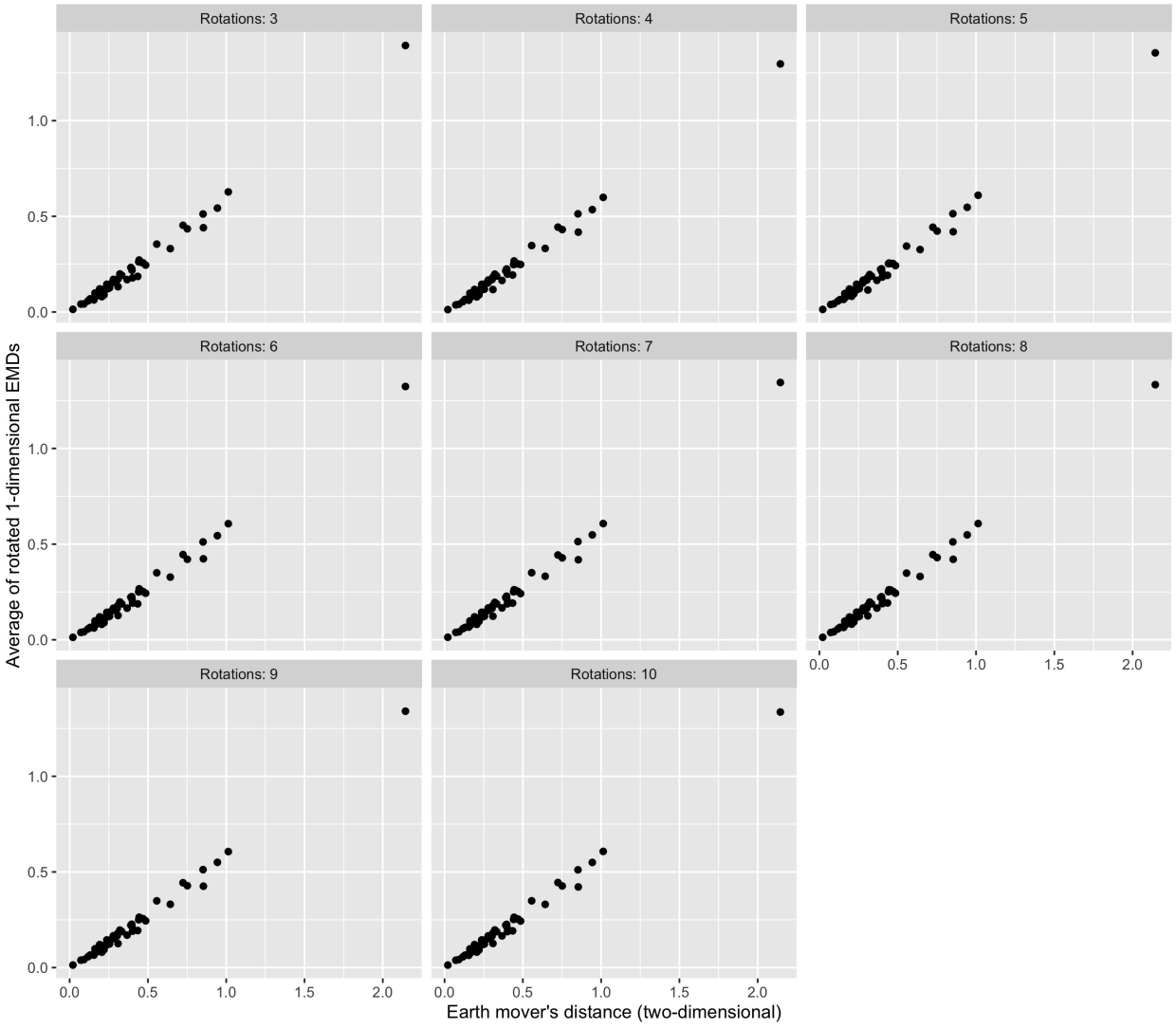


Figure A.2

There is no expectation that the two procedures would agree perfectly. For example, suppose to begin with that two distributions are identical; then the 2-dimensional EMD will be zero, as will the 1-dimensional EMD in each rotation, so the two measures will agree. If we then shift one distribution one unit to the east, the EMD will be approximately 1; the 1-dimensional CDF discrepancy will be 1 in the east-west direction, 0 in the north-south direction, and something in between in other directions (so that the mean will be between 0 and 1). The properties of the proposed proxy may require deeper investigation for other uses, but for the purpose of this paper we seek only to show that the proxy agrees closely with the two-dimensional EMD in the data we analyze. To show that it is the case, we compute the EMD and the proposed proxy (with a number of rotations ranging from 3 to 10), and compare the distribution of legislator birthplaces to the distribution of the population (both gridded) in 53 countries (all but the largest 10). The results plotted below show that the two measures agree very closely. In figure A.2, we show the scatter plots of the EMD

and its proxy for the 53 countries in the restricted sample across different parameters for the number of rotations. Figure A.3 shows how the correlation (in red) and the correlation of ranks (in blue) varies with the number of rotations. For this dataset, the correlation of ranks is slightly lower for lower numbers of rotations, but all correlations are well above .95, suggesting the proxy is valid for our purposes across all values tested.

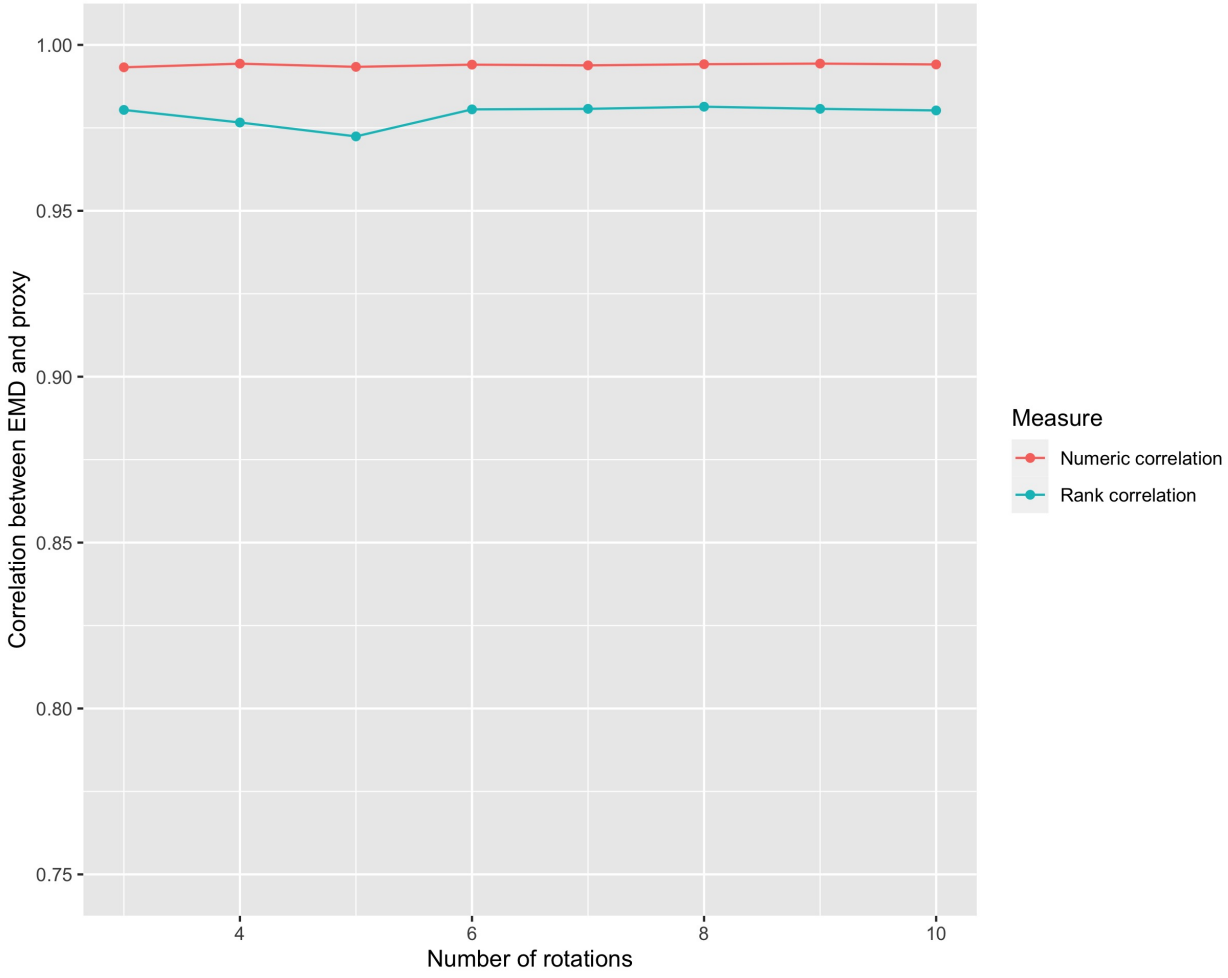
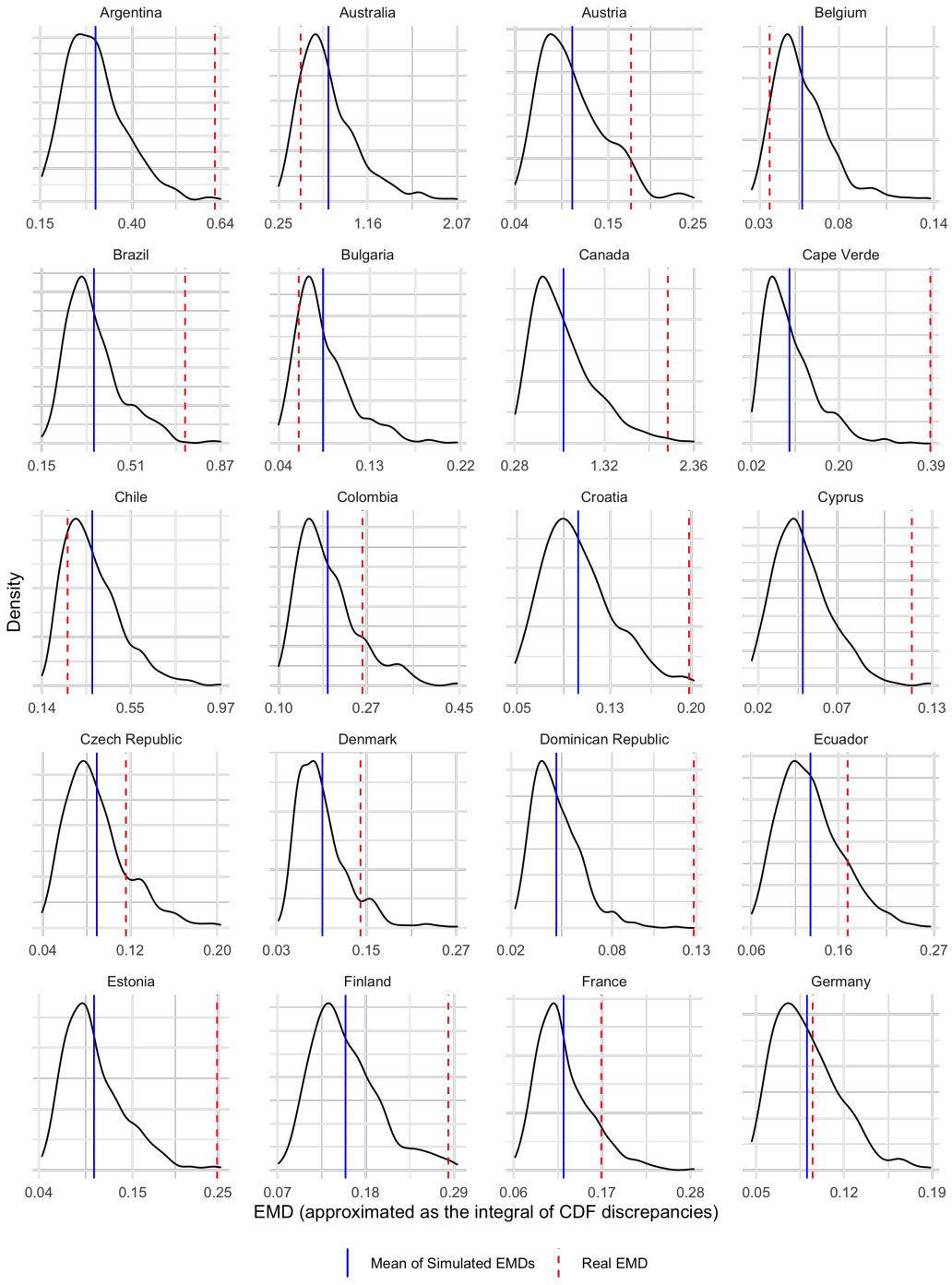
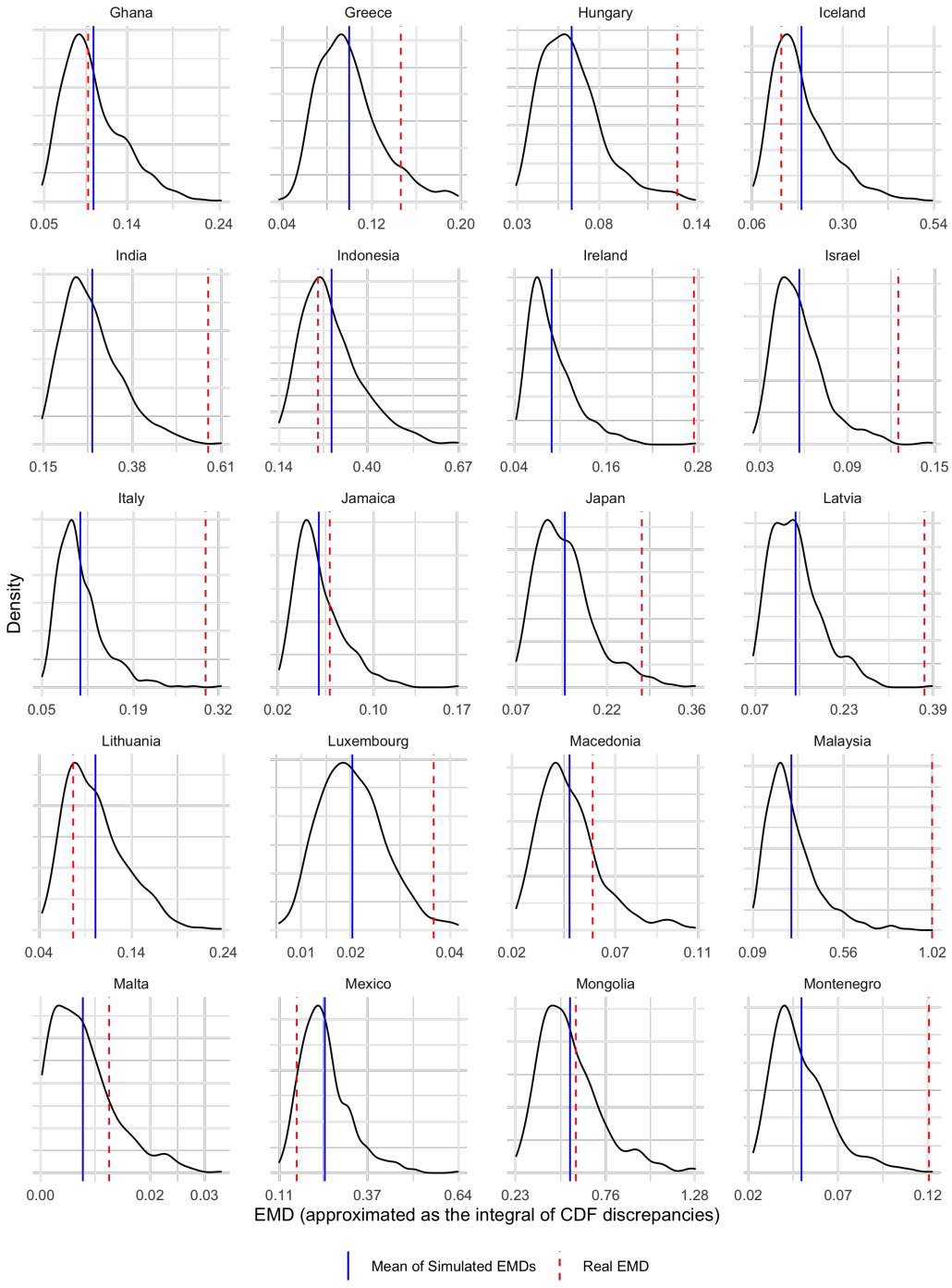


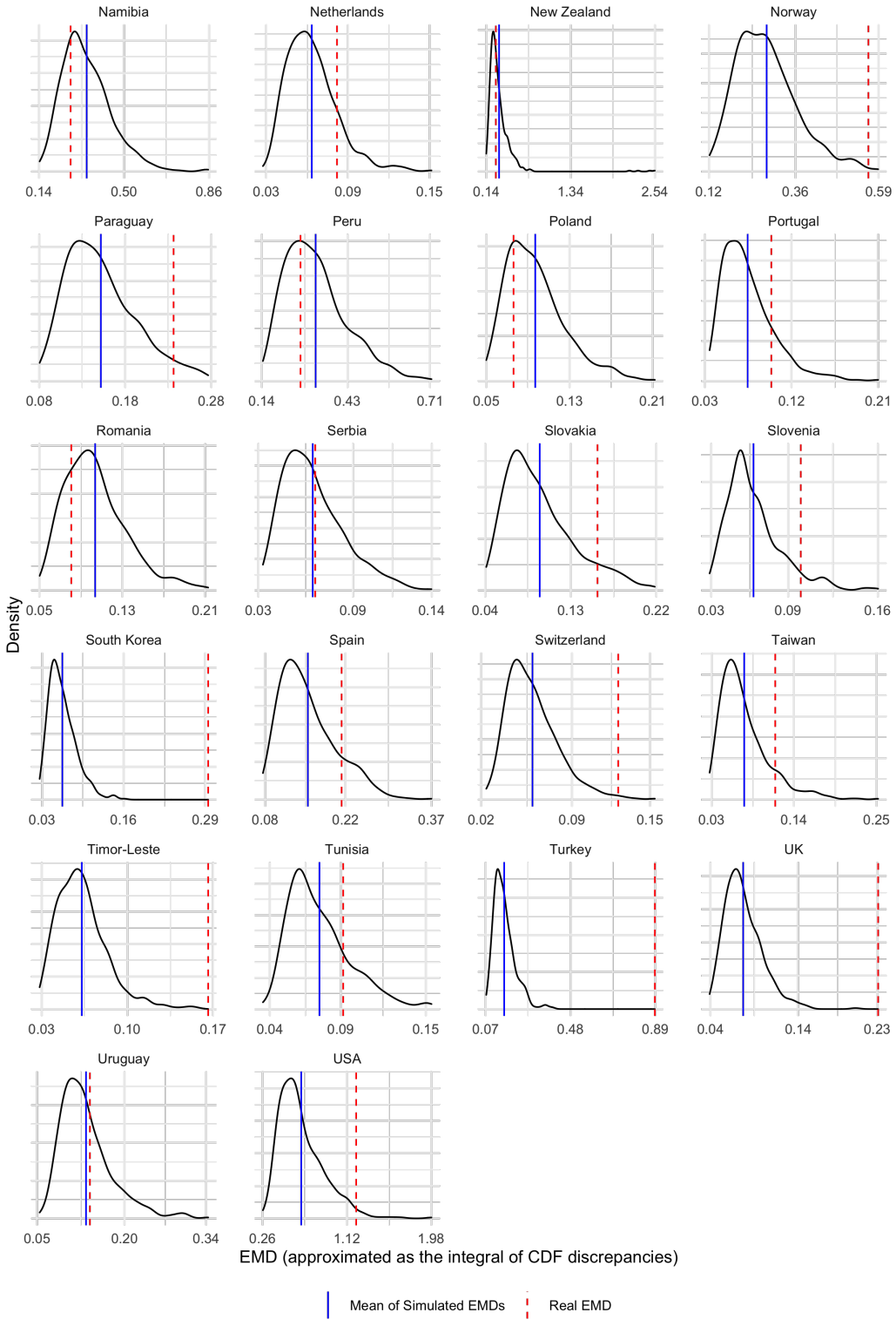
Figure A.3

B First Paper: Distribution of Discrepancy Estimates

Figure B.1: Density plot of EMD values across random draws of the assembly (2005 population benchmark). The value of the spatial discrepancy indicator for the actual parliament is plotted as a solid red line; the mean of the simulated EMDs is plotted as a solid blue line. Figure continues on following pages.




Ghana
0.05 0.14 0.24
Greece
0.04 0.12 0.20
Hungary
0.03 0.08 0.14
Iceland
0.06 0.30 0.54
India
0.15 0.38 0.61
Indonesia
0.14 0.40 0.67
Ireland
0.04 0.16 0.28
Israel
0.03 0.09 0.15
Italy
0.05 0.19 0.32
Jamaica
0.02 0.10 0.17
Japan
0.07 0.22 0.36
Latvia
0.07 0.23 0.39
Lithuania
0.04 0.14 0.24
Luxembourg
0.01 0.02 0.04
Macedonia
0.02 0.07 0.11
Malaysia
0.09 0.56 1.02
Malta
0.00 0.02 0.03
Mexico
0.11 0.37 0.64
Mongolia
0.23 0.76 1.28
Montenegro
0.02 0.07 0.12
Density
EMD (approximated as the integral of CDF discrepancies)
Mean of Simulated EMDs
Real EMD


Namibia
0.14
0.50
0.86
Netherlands
0.03
0.09
0.15
New Zealand
0.14
1.34
2.54
Norway
0.12
0.36
0.59
Paraguay
0.08
0.18
0.28
Peru
0.14
0.43
0.71
Poland
0.05
0.13
0.21
Portugal
0.03
0.12
0.21
Romania
0.05
0.13
0.21
Serbia
0.03
0.09
0.14
Slovakia
0.04
0.13
0.22
Slovenia
0.03
0.09
0.16
South Korea
0.03
0.16
0.29
Spain
0.08
0.22
0.37
Switzerland
0.02
0.09
0.15
Taiwan
0.03
0.14
0.25
Timor-Leste
0.03
0.10
0.17
Tunisia
0.04
0.09
0.15
Turkey
0.07
0.48
0.89
UK
0.04
0.14
0.23
Uruguay
0.05
0.20
0.34
USA
0.26
1.12
1.98
Density
EMD (approximated as the integral of CDF discrepancies)
Mean of Simulated EMDs
Real EMD

C First Paper: Robustness Checks (Cross-country analysis)

Table C.1: SURLI 2005 benchmark, categorical constituency structure variable: alternative controls

	Dependent variable:				
	SURLI (2005 benchmark)				
	(1)	(2)	(3)	(4)	(5)
Multi-Member	2.81** (1.17)	2.79** (1.19)	2.53** (1.10)	2.70** (1.12)	2.67** (1.13)
Single-Member	1.32 (1.28)	1.16 (1.29)	1.19 (1.33)	0.95 (1.34)	0.72 (1.35)
Preferential Voting	−1.73** (0.82)	−1.58* (0.84)	−1.64* (0.82)	−1.79** (0.82)	−1.65* (0.84)
Log(Median DM)	−0.40 (0.30)	−0.49 (0.32)			
Log(Mean DM)			−0.45 (0.31)	−0.46 (0.31)	−0.56* (0.33)
log(Population)	0.64** (0.32)	0.62* (0.35)	0.39 (0.32)	0.62* (0.32)	0.61* (0.35)
log(Land area)	−0.48* (0.26)	−0.63* (0.34)	−0.42 (0.25)	−0.47* (0.25)	−0.61* (0.34)
log(GDP pc)	0.43 (0.34)	0.45 (0.38)	1.08* (0.56)	0.43 (0.34)	0.45 (0.38)
Federalism	−0.95 (0.99)			−0.86 (0.99)	
Spatial Gini		1.78 (18.97)			1.39 (18.78)
Democracy score			−1.67 (1.09)		
Constant	−7.15 (5.65)	−5.40 (6.19)	−7.91 (5.50)	−6.66 (5.67)	−4.96 (6.18)
Observations	62	60	62	62	60
R^2	0.18	0.19	0.21	0.19	0.20
Adjusted R^2	0.06	0.06	0.09	0.07	0.07
Note:					*p<0.1; **p<0.05; ***p<0.01

Table C.2: SURLI legislator birthyear benchmark, categorical constituency structure variable: alternative controls

	Dependent variable:				
	SURLI (mean legislator birth year benchmark)				
	(1)	(2)	(3)	(4)	(5)
Multi-Member	1.66* (0.99)	1.63 (1.00)	1.56* (0.93)	1.62* (0.95)	1.56 (0.95)
Single-Member	1.64 (1.08)	1.96* (1.08)	2.27** (1.12)	1.68 (1.13)	2.02* (1.14)
Preferential Voting	−1.01 (0.69)	−0.77 (0.70)	−0.78 (0.69)	−0.99 (0.70)	−0.74 (0.71)
log(Median DM)	−0.001 (0.26)	−0.01 (0.27)			
log(Mean DM)			0.05 (0.26)	0.03 (0.26)	0.04 (0.28)
log(Population)	0.39 (0.27)	0.61** (0.30)	0.29 (0.27)	0.40 (0.27)	0.62** (0.30)
log(Land area)	−0.39* (0.22)	−0.57* (0.28)	−0.27 (0.21)	−0.38* (0.22)	−0.57** (0.28)
log(GDP pc)	−0.02 (0.29)	0.21 (0.32)	0.67 (0.47)	−0.02 (0.29)	0.21 (0.32)
Federalism	0.59 (0.84)			0.59 (0.84)	
Spatial Gini		14.63 (15.91)			15.33 (15.83)
Democracy score			−1.55* (0.91)		
Constant	−0.59 (4.77)	−4.80 (5.19)	−5.25 (4.62)	−0.66 (4.79)	−5.00 (5.21)
Observations	62	60	62	62	60
R^2	0.14	0.15	0.18	0.14	0.15
Adjusted R^2	0.01	0.02	0.06	0.01	0.02
Note:					*p<0.1; **p<0.05; ***p<0.01

Table C.3: SURLI 2005 benchmark, continuous constituency structure variable plus MXM dummy: alternative controls

	Dependent variable:				
	SURLI (2005 benchmark)				
	(1)	(2)	(3)	(4)	(5)
Share multi-member	1.06 (1.33)	1.10 (1.38)	0.94 (1.43)	1.36 (1.42)	1.49 (1.46)
Mixed	−2.03* (1.03)	−1.93* (1.03)	−1.85* (1.01)	−1.79* (1.03)	−1.65 (1.02)
Preferential Voting	−1.62* (0.82)	−1.44* (0.84)	−1.54* (0.82)	−1.69** (0.83)	−1.54* (0.84)
log(Median DM)	−0.35 (0.30)	−0.42 (0.32)			
log(Mean DM)			−0.41 (0.31)	−0.41 (0.32)	−0.51 (0.33)
log(Population)	0.64* (0.32)	0.63* (0.36)	0.37 (0.32)	0.63* (0.32)	0.62* (0.36)
log(Land area)	−0.48* (0.26)	−0.66* (0.34)	−0.42 (0.25)	−0.47* (0.26)	−0.64* (0.34)
log(GDP pc)	0.43 (0.34)	0.46 (0.38)	1.11* (0.56)	0.43 (0.34)	0.46 (0.38)
Federalism	−1.01 (0.99)			−0.94 (0.99)	
Spatial Gini		4.02 (18.91)			3.29 (18.77)
Democracy score			−1.76 (1.08)		
Constant	−5.52 (5.70)	−4.08 (6.24)	−6.52 (5.48)	−5.52 (5.68)	−4.14 (6.18)
Observations	62	60	62	62	60
R^2	0.18	0.18	0.21	0.18	0.19
Adjusted R^2	0.05	0.05	0.09	0.06	0.06
Note:					*p<0.1; **p<0.05; ***p<0.01

Table C.4: SURLI mean legislator birthyear benchmark, continuous constituency structure variable plus MXM dummy: alternative controls

	Dependent variable:				
	SURLI (mean legislator birth year benchmark)				
	(1)	(2)	(3)	(4)	(5)
Share multi-member	0.11 (1.12)	−0.15 (1.15)	−0.49 (1.20)	0.04 (1.19)	−0.26 (1.23)
Mixed	−1.64* (0.86)	−1.79** (0.86)	−1.92** (0.84)	−1.64* (0.86)	−1.79** (0.86)
Preferential Voting	−1.04 (0.69)	−0.82 (0.70)	−0.84 (0.69)	−1.02 (0.70)	−0.79 (0.71)
log(Median DM)	−0.01 (0.25)	−0.03 (0.26)			
log(Mean DM)			0.03 (0.26)	0.02 (0.27)	0.01 (0.28)
log(Population)	0.40 (0.27)	0.61** (0.30)	0.30 (0.27)	0.40 (0.27)	0.62** (0.30)
log(Land area)	−0.39* (0.22)	−0.56* (0.28)	−0.27 (0.21)	−0.38* (0.22)	−0.56* (0.28)
log(GDP pc)	−0.02 (0.29)	0.21 (0.32)	0.65 (0.47)	−0.02 (0.29)	0.21 (0.32)
Federalism	0.60 (0.83)			0.60 (0.83)	
Spatial Gini		13.93 (15.77)			14.58 (15.74)
Democracy score			−1.50 (0.91)		
Constant	0.92 (4.78)	−2.94 (5.20)	−3.10 (4.59)	0.91 (4.78)	−3.07 (5.18)
Observations	62	60	62	62	60
R^2	0.14	0.15	0.18	0.14	0.15
Adjusted R^2	0.01	0.02	0.05	0.01	0.02
Note:					*p<0.1; **p<0.05; ***p<0.01

Table C.5: Main model, excludes micro-countries (number of grids ≤ 20)

	Dependent variable:			
	SURLI (2005)		SURLI (mean legislator birth year)	
	(1)	(2)	(3)	(4)
Multi-Member	2.39** (1.18)		1.34 (0.98)	
Single-Member	1.81 (1.31)		2.35** (1.08)	
Share Multi-Member		0.19 (1.39)		−0.73 (1.15)
Mixed		−2.12** (1.03)		−1.88** (0.85)
Preferential Voting	−1.22 (0.87)	−1.10 (0.86)	−0.52 (0.72)	−0.60 (0.71)
log(Median DM)	−0.33 (0.31)	−0.29 (0.31)	0.09 (0.26)	0.05 (0.25)
log(Population)	0.38 (0.33)	0.36 (0.33)	0.29 (0.27)	0.30 (0.27)
log(Land area)	−0.42 (0.29)	−0.41 (0.29)	−0.20 (0.24)	−0.20 (0.24)
log(GDP pc)	1.45** (0.65)	1.49** (0.65)	1.00* (0.54)	0.96* (0.54)
Democracy score	−2.28* (1.22)	−2.39* (1.21)	−2.10** (1.01)	−2.01* (1.00)
Constant	−11.29* (6.28)	−9.35 (6.24)	−8.74* (5.21)	−6.38 (5.18)
Observations	58	58	58	58
R^2	0.23	0.23	0.21	0.20
Adjusted R^2	0.10	0.10	0.08	0.07
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C.6: Main model, excludes US

	Dependent variable:			
	SURLI (2005)		SURLI (mean legislator birth year)	
	(1)	(2)	(3)	(4)
Multi-Member	2.65** (1.16)		1.62 (0.97)	
Single-Member	1.64 (1.32)		2.11* (1.11)	
Share Multi-Member		0.56 (1.37)		−0.29 (1.15)
Mixed		−2.15** (1.04)		−1.87** (0.87)
Preferential Voting	−1.58* (0.82)	−1.47* (0.82)	−0.82 (0.69)	−0.87 (0.69)
log(Median DM)	−0.40 (0.30)	−0.35 (0.30)	−0.01 (0.25)	−0.03 (0.25)
log(Population)	0.39 (0.32)	0.37 (0.33)	0.29 (0.27)	0.30 (0.27)
log(Land area)	−0.42 (0.26)	−0.42 (0.26)	−0.28 (0.22)	−0.28 (0.22)
log(GDP pc)	1.13* (0.58)	1.16** (0.58)	0.65 (0.48)	0.64 (0.48)
Democracy score	−1.76 (1.10)	−1.85* (1.10)	−1.53 (0.92)	−1.49 (0.92)
Constant	−8.69 (5.70)	−6.82 (5.64)	−4.81 (4.78)	−2.80 (4.72)
Observations	61	61	61	61
R^2	0.21	0.20	0.17	0.17
Adjusted R^2	0.09	0.08	0.04	0.04
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C.7: Main model, assembly size used instead of logged population

	Dependent variable:			
	SURLI (2005 benchmark)		SURLI (mean legislator birth year)	
	(1)	(2)	(3)	(4)
Multi-Member	2.74** (1.12)		1.86** (0.91)	
Single-Member	1.66 (1.25)		2.27** (1.02)	
Share Multi-Member		0.57 (1.28)		−0.30 (1.04)
Mixed		−2.20** (0.99)		−2.08** (0.81)
Preferential Voting	−1.40* (0.80)	−1.27 (0.80)	−0.66 (0.65)	−0.68 (0.65)
log(Median DM)	−0.43 (0.29)	−0.36 (0.29)	−0.03 (0.24)	−0.04 (0.23)
Assembly Size	0.005* (0.002)	0.005* (0.002)	0.01** (0.002)	0.01** (0.002)
log(Land area)	−0.39* (0.20)	−0.40* (0.20)	−0.31* (0.16)	−0.31* (0.16)
log(GDP pc)	1.02* (0.55)	1.06* (0.55)	0.52 (0.45)	0.52 (0.44)
Democracy score	−1.75* (1.03)	−1.85* (1.02)	−1.41* (0.84)	−1.39* (0.83)
Constant	−2.94 (5.22)	−1.14 (5.18)	−0.11 (4.25)	2.15 (4.20)
Observations	62	62	62	62
R^2	0.24	0.23	0.25	0.25
Adjusted R^2	0.13	0.12	0.14	0.14
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C.8: Main model, controls for both assembly size and logged population

	Dependent variable:			
	SURLI (2005 benchmark)		SURLI (mean legislator birth year benchmark)	
	(1)	(2)	(3)	(4)
Multi-Member	2.72** (1.14)		1.75* (0.92)	
Single-Member	1.68 (1.26)		2.36** (1.02)	
Share Multi-Member		0.50 (1.33)		−0.55 (1.07)
Mixed		−2.21** (1.00)		−2.09** (0.81)
Preferential Voting	−1.38* (0.82)	−1.23 (0.82)	−0.55 (0.66)	−0.56 (0.66)
log(Median DM)	−0.43 (0.30)	−0.37 (0.29)	−0.05 (0.24)	−0.06 (0.23)
Assembly Size	0.01 (0.003)	0.01 (0.003)	0.01** (0.003)	0.01** (0.003)
log(Population)	−0.07 (0.44)	−0.10 (0.45)	−0.35 (0.36)	−0.35 (0.36)
log(Land Area)	−0.37 (0.26)	−0.36 (0.26)	−0.19 (0.21)	−0.19 (0.21)
log(GDP pc)	1.04* (0.56)	1.07* (0.56)	0.59 (0.45)	0.58 (0.45)
Democracy Score	−1.80 (1.08)	−1.91* (1.08)	−1.64* (0.87)	−1.63* (0.86)
Constant	−2.21 (6.81)	−0.17 (6.94)	3.24 (5.50)	5.73 (5.58)
Observations	62	62	62	62
R^2	0.24	0.23	0.27	0.27
Adjusted R^2	0.11	0.10	0.14	0.14
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C.9: Main model, control for presidentialism

	Dependent variable:			
	SURLI (2005 benchmark)		SURLI (mean legislator birth year)	
	(1)	(2)	(3)	(4)
Multi-Member	2.67** (1.15)		1.67* (0.95)	
Single-Member	1.17 (1.31)		1.74 (1.08)	
Share Multi-Member		0.91 (1.36)		−0.004 (1.11)
Mixed		−1.91* (1.02)		−1.70** (0.84)
Preferential Voting	−1.57* (0.81)	−1.42* (0.81)	−0.82 (0.67)	−0.83 (0.67)
log(Median DM)	−0.46 (0.30)	−0.39 (0.29)	−0.07 (0.25)	−0.08 (0.24)
log(Population)	0.29 (0.33)	0.27 (0.33)	0.17 (0.27)	0.18 (0.27)
log(Land area)	−0.34 (0.26)	−0.35 (0.26)	−0.17 (0.22)	−0.17 (0.22)
log(GDP pc)	0.99* (0.57)	1.04* (0.57)	0.54 (0.47)	0.54 (0.47)
Democracy Score	−1.92* (1.09)	−2.02* (1.10)	−1.77* (0.90)	−1.76* (0.90)
Presidential	−1.01 (0.81)	−0.91 (0.80)	−1.17* (0.66)	−1.18* (0.66)
Constant	−5.69 (5.84)	−4.40 (5.77)	−2.11 (4.82)	−0.41 (4.74)
Observations	62	62	62	62
R^2	0.23	0.22	0.22	0.22
Adjusted R^2	0.10	0.09	0.09	0.09
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C.10: Main model, with interaction term log(Magnitude) × Preferential Voting

	Dependent variable:			
	SURLI (2005 benchmark)		SURLI (mean legislator birth year benchmark)	
	(1)	(2)	(3)	(4)
Multi-Member	2.69** (1.16)		1.63 (0.98)	
Single-Member	1.39 (1.32)		2.20* (1.11)	
Share Multi-Member		0.80 (1.40)		−0.35 (1.17)
Mixed		−2.03* (1.03)		−1.92** (0.86)
Preferential Voting	−2.22 (1.48)	−1.97 (1.47)	−0.78 (1.25)	−0.89 (1.23)
log(median DM)	−0.51 (0.37)	−0.44 (0.36)	0.0003 (0.31)	−0.03 (0.30)
PV × log(med. DM)	0.31 (0.60)	0.25 (0.59)	−0.02 (0.50)	0.005 (0.50)
log(Population)	0.34 (0.33)	0.33 (0.34)	0.29 (0.28)	0.30 (0.28)
log(Land Area)	−0.40 (0.26)	−0.40 (0.26)	−0.28 (0.22)	−0.28 (0.22)
log(GDP pc)	1.09* (0.57)	1.13* (0.57)	0.68 (0.48)	0.66 (0.48)
Democracy score	−1.80 (1.10)	−1.88* (1.10)	−1.54 (0.93)	−1.51 (0.92)
Constant	−7.59 (5.70)	−6.02 (5.64)	−5.18 (4.79)	−3.06 (4.72)
Observations	62	62	62	62
R^2	0.21	0.21	0.18	0.18
Adjusted R^2	0.08	0.07	0.04	0.03
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C.11: Main model, Number of districts used instead of median district magnitude.

	Dependent variable:			
	SURLI (2005 benchmark)		SURLI (mean legislator birth year)	
	(1)	(2)	(3)	(4)
Multi-Member	2.61** (1.09)		1.72* (0.92)	
Single-Member	1.21 (1.29)		2.10* (1.09)	
Share Multi-Member		0.99 (1.35)		−0.18 (1.14)
Mixed		−1.90* (0.99)		−1.89** (0.84)
Preferential Voting	−1.58* (0.80)	−1.48* (0.80)	−0.86 (0.67)	−0.91 (0.67)
log(No. Districts)	0.52* (0.29)	0.48 (0.29)	0.08 (0.25)	0.10 (0.25)
log(Population)	0.18 (0.34)	0.18 (0.34)	0.26 (0.29)	0.26 (0.29)
log(Land area)	−0.40 (0.25)	−0.40 (0.25)	−0.28 (0.21)	−0.28 (0.21)
log(GDP pc)	1.06* (0.56)	1.10* (0.56)	0.68 (0.47)	0.66 (0.47)
Democracy score	−1.71 (1.07)	−1.81* (1.07)	−1.55* (0.91)	−1.51 (0.91)
Constant	−7.22 (5.49)	−5.83 (5.46)	−4.88 (4.65)	−2.91 (4.61)
Observations	62	62	62	62
R^2	0.23	0.22	0.18	0.18
Adjusted R^2	0.11	0.10	0.06	0.05
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C.12: Main model, control for countries with a residency requirement rule (USA, Argentina, Brazil, Chile, Taiwan, Ecuador), coded from Massicotte, Blais and Yoshinaka (2004).

	Dependent variable:			
	SURLI (2005 benchmark)		SURLI (mean legislator birth year)	
	(1)	(2)	(3)	(4)
Multi-Member	2.70** (1.16)		1.59 (0.97)	
Single-Member	1.53 (1.28)		2.21** (1.08)	
Share Multi-Member		0.71 (1.35)		−0.40 (1.13)
mixed		−2.10** (1.02)		−1.91** (0.85)
Preferential Voting	−1.54* (0.82)	−1.42* (0.82)	−0.85 (0.69)	−0.91 (0.69)
log(Median DM)	−0.43 (0.30)	−0.38 (0.30)	0.02 (0.25)	−0.01 (0.25)
log(Population)	0.40 (0.32)	0.38 (0.32)	0.28 (0.27)	0.29 (0.27)
log(Land area)	−0.40 (0.26)	−0.40 (0.26)	−0.30 (0.22)	−0.30 (0.22)
log(GDP pc)	1.18** (0.58)	1.20** (0.58)	0.62 (0.48)	0.61 (0.48)
Democracy score	−1.89* (1.12)	−1.98* (1.12)	−1.43 (0.94)	−1.39 (0.93)
Residency requirement	−0.82 (1.23)	−0.78 (1.24)	0.62 (1.03)	0.60 (1.03)
Constant	−9.33 (5.72)	−7.44 (5.75)	−4.36 (4.80)	−2.31 (4.80)
Observations	62	62	62	62
R^2	0.21	0.21	0.18	0.18
Adjusted R^2	0.08	0.07	0.04	0.04
Note:				*p<0.1; **p<0.05; ***p<0.01

D First Paper: Robustness Checks (Within-country Analysis)

Table D.1: Main models for UK MPs, probit link

	Dependent variable:	
	MP Born in Seat	
	(1)	(2)
Party margin in previous election	−0.579** (0.267)	−0.707** (0.294)
Constituency area	−0.0001 (0.0001)	−0.0001* (0.0001)
Labour	0.748*** (0.121)	1.414*** (0.421)
Lib Dem	0.235 (0.205)	1.183** (0.514)
Other	0.781*** (0.233)	1.254** (0.576)
SNP	0.443** (0.191)	1.086 (0.779)
2005	−0.018 (0.193)	0.445 (0.411)
2010	0.125 (0.176)	0.591 (0.374)
2015	0.401** (0.187)	0.980** (0.388)
2017	0.334 (0.204)	0.995** (0.428)
2019	0.495*** (0.186)	1.088*** (0.377)
By-election	0.121 (0.396)	0.849 (0.770)
Constant	−0.943*** (0.167)	−1.446*** (0.355)
Party × Election Interaction	No	Yes
Observations	864	864
Log Likelihood	−510.830	−497.280
Akaike Inf. Crit.	1,047.660	1,064.559
Note:		*p<0.1; **p<0.05; ***p<0.01

Table D.2: Main models for German SMD MPs, probit link

	Dependent variable:	
	MP Born in Seat (German SM tier)	
	(1)	(2)
Party margin in previous election	0.747* (0.416)	0.957** (0.479)
constituency_land_area	−0.0001** (0.0001)	−0.0001* (0.0001)
Other	0.218 (0.312)	4.630 (235.034)
SPD	0.048 (0.138)	−0.052 (0.344)
2002	0.303* (0.181)	0.235 (0.365)
2005	0.278 (0.207)	0.161 (0.375)
2009	0.445** (0.191)	0.440 (0.347)
2013	0.538** (0.216)	0.373 (0.357)
2017	0.172 (0.205)	−0.099 (0.352)
Constant	0.290* (0.176)	0.358 (0.327)
Party × Election Interaction	No	Yes
Observations	604	604
Log Likelihood	−365.340	−361.355
Akaike Inf. Crit.	750.680	760.710
Note:		*p<0.1; **p<0.05; ***p<0.01

Table D.3: UK models, no interaction controls, alternative specification of covariates

	Dependent variable:		
	MP born in seat (UK)		
	(1)	(2)	(3)
Party margin in previous election	−0.87* (0.45)	−0.90** (0.45)	−0.95** (0.45)
Log(Constituency area)	−0.18*** (0.06)		
Constituency area		−0.0002* (0.0001)	−0.0002 (0.0001)
Country[a]			
Northern Ireland		0.80 (1.01)	
Scotland		0.48 (0.32)	
Wales		0.37 (0.31)	
MP's Gender (male)			0.27 (0.17)
Party[b]			
Labour	1.06*** (0.21)	1.16*** (0.21)	1.31*** (0.21)
Lib Dem	0.27 (0.35)	0.33 (0.35)	0.43 (0.35)
Other	1.43*** (0.38)	0.65 (0.94)	1.34*** (0.38)
SNP	0.77** (0.31)	0.30 (0.43)	0.75** (0.31)
Election[c]			
2005	−0.09 (0.33)	0.02 (0.33)	0.02 (0.33)
2010	0.17 (0.30)	0.26 (0.30)	0.27 (0.30)
2015	0.58* (0.32)	0.74** (0.32)	0.76** (0.32)
2017	0.48 (0.34)	0.51 (0.34)	0.56 (0.35)
2019	0.74** (0.31)	0.87*** (0.32)	0.93*** (0.32)
By-election	0.07 (0.66)	0.21 (0.67)	0.19 (0.66)
Constant	−0.64 (0.43)	−1.62*** (0.29)	−1.82*** (0.34)
Party × Election Interaction	No	No	No
Observations	864	864	864
Log Likelihood	−506.54	−508.15	−508.66
Akaike Inf. Crit.	1,039.09	1,048.30	1,045.33

[a] = ref. cat. England, [b] = ref. cat. Conservative, [c] = ref. cat. 2001.
*p<0.1; **p<0.05; ***p<0.01

Table D.4: UK models, interaction controls, alternative specification of covariates

	Dependent variable:		
	MP born in seat (UK)		
	(1)	(2)	(3)
Party margin in previous election	−1.06** (0.49)	−1.04** (0.50)	−1.14** (0.49)
log(Constituency area)	−0.19*** (0.06)		
Constituency area		−0.0002* (0.0001)	−0.0002* (0.0001)
Country[a]			
Northern Ireland		0.76 (1.16)	
Scotland		0.49 (0.34)	
Wales		0.33 (0.31)	
MP's gender (male)			0.22 (0.17)
Party[b]			
Labour	2.29*** (0.82)	2.26*** (0.83)	2.49*** (0.82)
Lib Dem	2.01** (0.96)	1.96** (0.96)	2.09** (0.96)
Other	2.37** (1.04)	1.61 (1.39)	2.31** (1.04)
SNP	2.06 (1.37)	1.53 (1.42)	2.02 (1.38)
Election[c]			
2005	0.81 (0.83)	0.86 (0.83)	0.88 (0.83)
2010	1.09 (0.76)	1.13 (0.76)	1.17 (0.76)
2015	1.74** (0.78)	1.77** (0.78)	1.84** (0.78)
2017	1.65* (0.84)	1.48* (0.85)	1.69** (0.84)
2019	1.91** (0.76)	1.94** (0.76)	2.02*** (0.76)
By-election	1.31 (1.38)	1.58 (1.37)	1.62 (1.37)
Constant	−1.59** (0.80)	−2.54*** (0.73)	−2.74*** (0.75)
Party × Election Interaction	Yes	Yes	Yes
Observations	864	864	864
Log Likelihood	−493.07	−495.12	−495.77
Akaike Inf. Crit.	1,056.14	1,066.24	1,063.55

[a] = ref. cat. England, [b] = ref. cat. Conservative, [c] = ref. cat. 2001.
*p<0.1; **p<0.05; ***p<0.01

Table D.5: German SM tier models, no interaction controls, alternative specification of covariates

	Dependent variable:		
	MP born in seat (German SM tier)		
	(1)	(2)	(3)
Party margin in previous election	1.19* (0.69)	1.07 (0.71)	1.21* (0.69)
log(Constituency area)	−0.04 (0.07)		
Constituency area		−0.0001 (0.0001)	−0.0002** (0.0001)
Region (West)		0.26 (0.25)	
MP's gender (male)			0.10 (0.20)
Party[a]			
Others	0.34 (0.52)	0.50 (0.54)	0.41 (0.53)
SPD	0.14 (0.23)	0.11 (0.23)	0.10 (0.23)
Election[b]			
2002	0.54* (0.30)	0.50* (0.30)	0.50* (0.30)
2005	0.48 (0.34)	0.46 (0.34)	0.45 (0.34)
2009	0.73** (0.32)	0.73** (0.32)	0.74** (0.32)
2013	0.89** (0.36)	0.91** (0.37)	0.89** (0.37)
2017	0.32 (0.33)	0.29 (0.34)	0.29 (0.34)
Constant	0.46 (0.57)	0.21 (0.38)	0.39 (0.33)
Party × Election Interaction	No	No	No
Observations	604	604	604
Log Likelihood	−367.36	−364.78	−365.21
Akaike Inf. Crit.	754.73	751.56	752.41

[a] = ref. cat. CDU/CSU, [b] = ref. cat. 1998; *p<0.1; **p<0.05; ***p<0.01

Table D.6: German SM tier models, interaction controls, alternative specification of covariates

	Dependent variable:		
	MP born in seat (German SM tier)		
	(1)	(2)	(3)
Party margin in previous election	1.59** (0.81)	1.53* (0.84)	1.64** (0.81)
log(Constituency area)	−0.01 (0.08)		
Constituency area		−0.0001 (0.0001)	−0.0002* (0.0001)
Region (West)		0.17 (0.25)	
MP's gender (male)			0.08 (0.20)
Others	14.11 (882.74)	13.94 (882.74)	14.01 (882.74)
SPD	−0.05 (0.56)	−0.06 (0.56)	−0.05 (0.57)
2002	0.39 (0.60)	0.38 (0.60)	0.40 (0.60)
2005	0.26 (0.62)	0.25 (0.62)	0.28 (0.62)
2009 election	0.72 (0.57)	0.73 (0.57)	0.75 (0.57)
2013	0.57 (0.59)	0.64 (0.59)	0.63 (0.59)
2017	−0.15 (0.57)	−0.16 (0.57)	−0.15 (0.58)
Party × Election Interaction	Yes	Yes	Yes
Observations	604	604	604
Log Likelihood	−362.78	−360.98	−361.12
Akaike Inf. Crit.	763.57	761.96	762.25

Note: *p<0.1; **p<0.05; ***p<0.01

Table D.7

E Second Paper: Robustness Checks

Table E.1: Main Model (logistic regression), disaggregated parliamentary/party position variables

	Dependent variable:		
	Candidacy (1 = Ran for Federal Election)		
	(1)	(2)	(3)
Constant	-4.37^{***} (1.13)	-3.47^{***} (1.26)	-5.25^{***} (1.37)
List PR	0.48^{***} (0.13)	0.36^{**} (0.15)	0.52^{***} (0.13)
Party Leader	0.28^{*} (0.17)	0.29 (0.18)	0.28 (0.17)
Deputy Leader/Chief Whip	0.08 (0.21)	0.08 (0.21)	0.07 (0.21)
Committee Chair	0.18 (0.23)	0.18 (0.23)	0.18 (0.23)
Speaker/Deputy Speaker	0.11 (0.12)	0.12 (0.12)	0.10 (0.13)
Executive Position	0.32 (0.26)	0.35 (0.27)	0.32 (0.26)
Party in State Govt.	-0.30^{***} (0.10)	-0.23^{**} (0.11)	-0.26^{**} (0.11)
Age	0.12^{***} (0.04)	0.12^{***} (0.04)	0.12^{***} (0.04)
Age^2	-0.001^{***} (0.0004)	-0.002^{***} (0.0004)	-0.001^{***} (0.0004)
Seniority	0.37^{***} (0.09)	0.37^{***} (0.10)	0.40^{***} (0.09)
Seniority^2	-0.03^{**} (0.01)	-0.03^{**} (0.01)	-0.03^{***} (0.01)
Preferential Vote	0.14 (0.19)	0.16 (0.22)	0.14 (0.20)
Time to Next State Election	-0.15^{***} (0.04)	-0.14^{***} (0.04)	-0.14^{***} (0.04)
Male	-0.05 (0.10)	-0.06 (0.10)	-0.05 (0.10)
Nagelkerke pseudo R^2	0.157	0.183	0.174
Observations	14305	14305	14305
State, Party, Election Controls	✓	✓	✓
State × Party Controls		✓	
Party × Election Controls			✓
Only Mixed-Member			
	(4)	(5)	(6)
Constant	-5.50^{***} (1.40)	-4.84^{***} (1.52)	-6.83^{***} (1.64)
List PR	0.42^{***} (0.13)	0.36^{**} (0.15)	0.49^{***} (0.14)
Party Leader	0.34^{*} (0.19)	0.36^{*} (0.20)	0.33 (0.20)
Deputy Leader/Chief Whip	-0.39 (0.27)	-0.41 (0.27)	-0.42 (0.27)
Committee Chair	0.53^{*} (0.29)	0.54^{*} (0.30)	0.53^{*} (0.29)
Speaker/Deputy Speaker	0.15 (0.15)	0.15 (0.15)	0.12 (0.15)
Executive Position	0.22 (0.28)	0.25 (0.29)	0.19 (0.28)
Party in State Govt.	-0.42^{***} (0.12)	-0.38^{***} (0.14)	-0.31^{**} (0.13)
Age	0.16^{***} (0.05)	0.18^{***} (0.05)	0.17^{***} (0.05)
Age^2	-0.002^{***} (0.001)	-0.002^{***} (0.001)	-0.002^{***} (0.001)
Seniority	0.36^{***} (0.12)	0.32^{***} (0.12)	0.39^{***} (0.12)
Seniority^2	-0.02 (0.02)	-0.01 (0.02)	-0.02 (0.02)
Preferential Vote	-0.03 (0.52)	0.38 (0.75)	0.05 (0.55)
Time to Next State Election	-0.17^{***} (0.04)	-0.16^{***} (0.04)	-0.15^{***} (0.04)
Male	0.07 (0.11)	0.07 (0.12)	0.07 (0.12)
Nagelkerke pseudo R^2	0.162	0.189	0.184
Observations	12387	12387	12387
State, Party, Election Controls	✓	✓	✓
State × Party Controls		✓	
Party × Election Controls			✓
Only Mixed-Member	✓	✓	✓
Note:			*p<0.1; **p<0.05; ***p<0.01

Table E.2: Main model (multinomial regression), disaggregated party/legislative position variables.

	Dependent variable:			
	Candidacy Type (Ref.: *Didn't Run*)			
	(1)		(2)	
	Insecure	*Secure*	*Insecure*	*Secure*
Constant	-2.98^{**}	-14.22^{***}	-3.96^{***}	-12.77^{***}
	(1.22)	(3.28)	(1.51)	(3.45)
List PR	0.81^{***}	0.32^{*}	0.61^{***}	0.30
	(0.18)	(0.19)	(0.19)	(0.20)
Party Leader	0.14	0.56^{**}	0.09	0.73^{**}
	(0.21)	(0.28)	(0.25)	(0.29)
Deputy Leader/Chief Whip	0.30	-0.59	-0.18	-1.19
	(0.23)	(0.52)	(0.29)	(0.72)
Committee Chair	-0.22	1.11^{**}	0.07	1.61^{**}
	(0.26)	(0.54)	(0.33)	(0.75)
Speaker/Deputy Speaker	0.16	-0.003	0.24	0.01
	(0.15)	(0.22)	(0.18)	(0.24)
Executive Position	-1.01	0.55^{*}	-0.85	0.47
	(0.72)	(0.30)	(0.73)	(0.32)
Party in State Govt.	-0.33^{***}	-0.15	-0.48^{***}	-0.24
	(0.12)	(0.16)	(0.16)	(0.18)
Age	0.07	0.37^{***}	0.10^{*} 0.32^{***}	
	(0.04)	(0.11)	(0.06)	(0.11)
Age^2	-0.001^{**}	-0.004^{***}	-0.001^{**}	-0.004^{***}
	(0.0005)	(0.001)	(0.001)	(0.001)
Seniority	0.40^{**}	0.60^{***}	0.61^{**}	0.53^{***}
	(0.17)	(0.15)	(0.28)	(0.17)
Seniority^2	-0.06^{**}	-0.03^{*}	-0.09^{*}	-0.02
	(0.03)	(0.02)	(0.05)	(0.02)
Preferential Voting	0.07	0.46	0.39	0.24
	(0.23)	(0.38)	(1.08)	(0.65)
Time to Next State Election	-0.08^{*}	-0.24^{***}	-0.10^{**}	-0.25^{***}
	(0.05)	(0.06)	(0.05)	(0.06)
Male	-0.12	0.08	0.11	-0.04
	(0.12)	(0.17)	(0.15)	(0.18)
Observations	14305		12387	
State, Election, Party FE	✓		✓	
Only Mixed-Member			✓	
Note:				*p<0.1; **p<0.05; ***p<0.01

F Second Paper: Alternative Specification of Dependent Variable

Table F.1: Multinomial regression: candidacy type coded according to outcome rather than prior security

	Dependent variable:			
	Candidacy Outcome (Ref.: *Didn't Run*)			
	(1)		(2)	
	Unsuccessful	*Successful*	*Unsuccessful*	*Successful*
Constant	-3.49^{***}	-10.52^{***}	-4.20^{***}	-11.29^{***}
	(1.20)	(3.09)	(1.44)	(3.48)
List PR	0.76^{***}	0.23	0.62^{***}	0.22
	(0.17)	(0.20)	(0.18)	(0.20)
Executive Position	-0.09	0.33	-0.01	0.21
	(0.47)	(0.32)	(0.47)	(0.35)
Legislative Position	0.03	-0.05	0.09	-0.13
	(0.14)	(0.17)	(0.16)	(0.19)
Party Position	0.20	0.28	-0.0001	0.24
	(0.14)	(0.18)	(0.17)	(0.20)
Party in State Govt.	-0.26^{**}	-0.31^{*}	-0.35^{**}	-0.43^{**}
	(0.12)	(0.16)	(0.15)	(0.18)
Age	0.08^{*}	0.27^{***}	0.11^{**}	0.31^{**}
	(0.04)	(0.10)	(0.05)	(0.12)
Age^2	-0.001^{**}	-0.003^{***}	-0.001^{**}	-0.004^{***}
	(0.0004)	(0.001)	(0.001)	(0.001)
Seniority	0.27^{**}	0.66^{***}	0.31	0.61^{***}
	(0.13)	(0.15)	(0.19)	(0.18)
Seniority^2	-0.03^{*}	-0.04^{**}	-0.03	-0.03
	(0.02)	(0.02)	(0.03)	(0.02)
Preferential Voting	0.12	0.30	0.22	-0.06
	(0.23)	(0.36)	(0.83)	(0.71)
Time to Next State Election	-0.07	-0.24^{***}	-0.10^{*}	-0.24^{***}
	(0.05)	(0.05)	(0.06)	(0.05)
Male	-0.02	-0.15	0.20	-0.19
	(0.12)	(0.15)	(0.15)	(0.17)
Observations	14305		12387	
State, Election, Party Controls	✓		✓	
Only Mixed-Member			✓	
Note:				*p<0.1; **p<0.05; ***p<0.01

G Third Paper: Within-District relationship between s, v_1 and N_c

In section 5.2.2, I motivated the assumption that v_1 is endogenous to expectations of seat gains by illustrating how the distribution of preference shares in a Polish district is 'flatter' for more successful parties and 'steeper' for smaller ones. The empirical plausibility of this assumption can be more rigorously assessed by considering the wider universe of districts considered in the analysis (see footnote 8). Tables G.1 and G.2 report the results of OLS regressions *with district-election fixed effects*, where list-level values of v_1 and N_c are regressed on s (the raw number of seats gained) and on $\frac{s}{M}$, the share of seats gained. Linear and log-log specifications are presented. Consistently with the assumption, larger list have lower values of v_1 and higher values of N_c. The coefficients are virtually identical when controlling for the number of candidates fielded c (tables G.3 and G.4), confirming that the relationship is largely independent from parties' nominating behaviour.

Table G.1: District-election fixed effects model: relationship between the number of seats and intra-party quantities.

	Dependent variable:			
	v_1	N_c	$\log(v_1)$	$\log(N_c)$
	(1)	(2)	(3)	(4)
s	-0.007^{***}	0.295^{***}		
	(0.001)	(0.017)		
$\log(s)$			-0.161^{***}	0.151^{***}
			(0.009)	(0.010)
District-Election F.E.	✓	✓	✓	✓
Observations	2,617	2,617	2,617	2,617
R^2	0.051	0.127	0.128	0.103
Adjusted R^2	−0.201	−0.105	−0.103	−0.136
F Statistic (df = 1; 2067)	110.923^{***}	300.391^{***}	304.531^{***}	236.575^{***}
Note:				*p<0.1; **p<0.05; ***p<0.01

Table G.2: District-election fixed effects model: relationship between the share of seats and intra-party quantities.

	Dependent variable:			
	v_1	N_c	$\log(v_1)$	$\log(N_c)$
	(1)	(2)	(3)	(4)
s/M	-0.200^{***}	6.004^{***}		
	(0.015)	(0.376)		
$\log(s/M)$			-0.161^{***}	0.151^{***}
			(0.009)	(0.010)
District-Election F.E.	✓	✓	✓	✓
Observations	2,617	2,617	2,617	2,617
R^2	0.077	0.110	0.128	0.103
Adjusted R^2	-0.168	-0.126	-0.103	-0.136
F Statistic (df = 1; 2067)	171.882^{***}	255.481^{***}	304.531^{***}	236.575^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

Table G.3: District-election fixed effects model: relationship between the share of seats and intra-party quantities, controlling for list length.

	Dependent variable:			
	v_1	N_c	$\log(v_1)$	$\log(N_c)$
	(1)	(2)	(3)	(4)
s	-0.007^{***}	0.284^{***}		
	(0.001)	(0.017)		
c	-0.013^{***}	0.196^{***}		
	(0.001)	(0.031)		
$\log(s)$			-0.153^{***}	0.139^{***}
			(0.009)	(0.010)
$\log(c)$			-0.558^{***}	0.816^{***}
			(0.069)	(0.072)
District-Election F.E.	✓	✓	✓	✓
Observations	2,617	2,617	2,617	2,617
R^2	0.096	0.144	0.155	0.155
Adjusted R^2	−0.145	−0.084	−0.070	−0.070
F Statistic (df = 2; 2066)	109.844^{***}	173.252^{***}	189.714^{***}	189.377^{***}
Note:				*p<0.1; **p<0.05; ***p<0.01

Table G.4: District-election fixed effects model: relationship between the share of seats and intra-party quantities, controlling for list length.

	Dependent variable:			
	v_1	N_c	$\log(v_1)$	$\log(N_c)$
	(1)	(2)	(3)	(4)
s/M	-0.174^{***}	5.614^{***}		
	(0.015)	(0.377)		
c/M	-0.317^{***}	4.757^{***}		
	(0.030)	(0.744)		
$\log(s/M)$			-0.153^{***}	0.139^{***}
			(0.009)	(0.010)
$\log(c/M)$			-0.558^{***}	0.816^{***}
			(0.069)	(0.072)
District-Election F.E.	✓	✓	✓	✓
Observations	2,617	2,617	2,617	2,617
R^2	0.125	0.127	0.155	0.155
Adjusted R^2	-0.108	-0.105	-0.070	-0.070
F Statistic (df = 2; 2066)	147.542^{***}	150.622^{***}	189.714^{***}	189.377^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

H Third Paper: bias analysis under different model specifications

Throughout the third paper, I have noted a series of possible concerns with the validity of the model tests and assumptions. In this section, the analysis in section 5.6.1 is replicated with some tweaks of the theoretical assumptions, the modelling choices and the definition of the sample. The subsections below detail the possible alternatives to the choices behind the main results, and present the values of the slope $\hat{k}$ and the bias $\frac{|\beta-\hat{k}|}{se}$ when the alternative choices are implemented. To be sure, most of these are not, strictly speaking, 'robustness checks': they are tests of different models or of different dimensions of model performance. Future developments in formal analysis of PLPR may make the case for their usage over the one presented.

H.1 Alternative Specification of c^*

In section 5.2.2 footnote 10, it was noted that the list-level (and the related district-level) formulas cannot predict the values of v_1 (and the related value of N_c) for non-seat-winning parties. This is because of the assumption that the number of pertinent vote-earning candidates is endogenous to the expectations of the number of seats a list will win, and the further assumption that parties are 'correct' on average about their performance, so that $E(s) = s$, and $c^* = (sc)^{\frac{1}{2}}$. But what of lists that gain no seats? I suggested a possible avenue to generalise the model to lists where $s = 0$ by setting $c^* = [(s+1)c]^{\frac{1}{2}}$. This model effectively formalises a potentially realistic assumption that, even in the least-competitive scenario, there will be at least one non-elected candidate who had a shot at election. Furthermore, it is appealing because c^* retains the 'plus one' element of its inter-party analogue, the number of pertinent vote-earning parties, which is – both at district level and nationwide – modelled as $N_s + 1$ (Shugart and Taagepera, 2017, 128; see also the close cognate notion of 'viable candidates' modelled as $M + 1$ in Cox, 1997).

The deriving list-level equations from this tweak to the assumptions are $v_1 = [(s+1)cp]^{-\frac{1}{4}}$ and $N_c = [(s+1)cp]^{-\frac{3}{8}}$. The district-level equations are $v_1 = [prM(M^{\frac{3}{8}}+1)]^{-\frac{1}{4}}$ and $N_c = [prM(M^{\frac{3}{8}}+1)]^{\frac{3}{8}}$. Note that v_s remains unchanged as (1) algebraically, the s in the formula is derived from the average of seat-winning candidate shares, not from v_1, and (2) conceptually, it is illogical to derive v_s when $s = 0$, i.e. there is no 'last-eligible' candidate in a list that gains zero seat. Relatedly, a potential drawback of this model specification is that the predictions for v_1 and v_s when $s = 1$ differ in expectation. The table below reports the observed slopes β and the bias relative to the expected values $\hat{k}$ for the modified list- and district-level models and the two dependent variables of interest, tested over the same sample as the main analysis. Of course, ideally we would want to test this model on *all* lists, but this is not possible with the current data, which were only collected for seat-winning lists.

Table H.1: Bias diagnostics under different specifications of c^*.

	List-level models		*District-level models*	
	v_1	N_c	v_1	N_c
base X	$(s+1)cp$	$(s+1)cp$	$prM(M^{\frac{3}{8}}+1)$	$prM(M^{\frac{3}{8}}+1)$
expected slope ($\hat{k}$)	-0.250	0.375	-0.250	0.375
observed slope (β)	-0.245	0.365	-0.240	0.358
$\frac{\lvert\beta-\hat{k}\rvert}{se}$	1.592	2.508	3.274	4.237

H.2 Embeddedness-adjusted district-level model

In section 5.3.2 footnote 16, I noted that Shugart and Taagepera (2017) revised their expected seat share for the first party in a district σ_1' to account for the 'embeddedness' of a district in the wider political system. They do so by introducing a k parameter which formalises the intuition that district-level competitiveness increases when a district is part of a larger political system. Consider for instance Barbados and the UK: both have a single-member seat plurality system, but the former has an assembly size S of 30 and the latter has an assembly size of 650. The latter will therefore produce parties that have only realistic chances in some districts, but nonetheless run throughout the country. This will result in higher inter-party competition in districts embedded in larger systems than districts embedded in smaller ones: for instance, the effective number of electoral parties in a British district is close to 3, while in Barbados is almost exactly 2. In multi-member seats, this 'embeddedness' effect is consequential because it means that as the ratio between magnitude and assembly size decreases, the number of small parties expected to win seats increases, and therefore – for instance – the expected share of seats for the first party is smaller.

Under this revised derivation, the expected value of σ_1' is $M^{-\frac{k}{2}}$. The value of k is district specific and set to equal $k = 0.5 + \frac{0.2076\ log_{10}(\frac{S}{M})}{M^{\frac{1}{4}}}$. In practice, k tends to 1 as M becomes infinitesimally small relative to the assembly size, and is exactly 0.5 when $M = S$. Details of its derivation are in Shugart and Taagepera (2017, pp. 174-177). For our purposes, what matters is that from the new formula for σ_1' it follows that the expected number of seats for the first party is $M \times s_1 = M^{1-\frac{k}{2}}$, and the expected number of seats for any party is $M^{\frac{2-k}{4}}$. As argued in section 5.3.2, this is the value of s expressed as a function of institutional variable, so that the resulting embeddedness-adjust district-level predictions are:

$$v_1 = \left(prM^{\frac{6-k}{4}}\right)^{-\frac{1}{4}} \qquad N_c = \left(prM^{\frac{6-k}{4}}\right)^{\frac{3}{8}} \qquad v_s = \left(prM^{3-k}\right)^{-\frac{1}{4}}$$

However, as discussed in section 5.7, I think that there is a more significant 'embeddedness' effect of relevance for intra-party competition – i.e. the fact that lists including 'big fish' politicians become more common as $\frac{M}{S}$ decreases – that is not captured by k. Thus, for the sake of simplicity, I chose to present a model that does not take into account the (presumably much smaller) bias engendered by the fact that as $\frac{M}{S}$ increases, more parties compete, and therefore individual parties have lower seat expectations. It is clear that the ratio between district magnitude and assembly size matters, but the way in which the embeddedness parameter k operationalises this effect is insufficient. In any case, the table below reproduces the model bias analysis using these revised predictions of the values of the dependent variables of interest. Note that, confusingly, $\hat{k}$ (k-hat) refers to the expected intercept of the whole base, while the k noted as the dependent variable X refers to the embeddedness parameter.

Table H.2: Bias diagnostics of embeddedness-adjusted model.

	District-level models		
	v_1	N_c	v_s
base X	$prM^{\frac{6-k}{4}}$	$prM^{\frac{6-k}{4}}$	prM^{3-k}
expected slope ($\hat{k}$)	-0.250	0.375	-0.250
observed slope (β)	-0.258	0.385	-0.248
$\frac{\lvert\beta-\hat{k}\rvert}{se}$	2.134	2.261	0.707

H.3 Symmetric Regression

As noted in section 5.5.1 footnote 19, in early tests of logical models, Taagepera and Shugart (Taagepera, 2007; Shugart, Bergman and Watt, 2013) occasionally employed 'symmetric regression', an estimator distinct from OLS insofar as it accounts for error-in-variables on the independent variable (see Taagepera, 2008, pp. 154-175). In a univariate regression, this is equivalent to a Simple Major Axis regression (Kimura, 1992). The estimator has enjoyed virtually no take-up in the wider literature, and presents econometric issues in generalising beyond two dimensions and computing clustered standard errors. In more recent work, Taagepera and Shugart themselves reverted back to log-log OLS to test both inter- (Li and Shugart, 2016; Shugart and Taagepera, 2017, pp. 109–113) and intra-party (Shugart and Taagepera, 2017, pp. 215–235) models of electoral systems. The table below presents model slopes and associated measures of bias for the symmetric regression models, computed on unclustered standard errors as per SBW replication files.

Table H.3: Bias diagnostics, symmetric regression used instead of OLS

	List-level			*List-level*			*SBW*	
	v_1	N_c	v_s	v_1	N_c	v_s	v_1	v_s
base X	scp	scp	s^4cp	$prM^{\frac{11}{8}}$	$prM^{\frac{11}{8}}$	$prM^{\frac{5}{2}}$	c	c
exp. slope	-0.250	0.375	-0.250	-0.250	0.375	-0.250	-0.500	-0.750
obs. slope	-0.278	0.407	-0.268	-0.268	0.392	-0.256	-0.459	-0.705
$\frac{\lvert\beta-\hat{k}\rvert}{se}$	5.460	4.110	3.799	3.590	2.547	0.373	8.603	3.711

H.4 Drop-One-Country Subsamples

As discussed in section 5.5.1 footnote 19, I repeated the model bias analysis on reduced samples obtained dropping alternately one of the nine countries under consideration to assuage concerns that the results are driven by country-specific factors. Sample dependency is not a major concern for the v_s and N_c models: especially in their district-level version they produce consistently coefficients close to the expectation $\hat{k}$. The models for v_s – which, as noted in section 5.3.1 rest on shakier theoretical ground – do however present on occasion fairly large values of the bias indicator.

		List-level			*District-level*		
Country Dropped	variable	v_1	N_c	v_s	v_1	N_c	v_s
	base X	scp	scp	s^4cp	$prM^{\frac{11}{8}}$	$prM^{\frac{11}{8}}$	$prM^{\frac{5}{2}}$
	exp. slope $\hat{k}$	-0.250	0.375	-0.250	-0.250	0.375	-0.250
Belgium	obs. slope	-0.276	0.407	-0.265	-0.264	0.392	-0.238
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	5.668	5.508	7.307	3.650	3.361	3.981
Cyprus	obs. slope	-0.261	0.389	-0.262	-0.252	0.378	-0.237
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	2.621	2.773	6.721	0.521	0.563	4.500
Czechia	obs. slope	-0.254	0.381	-0.259	-0.245	0.369	-0.234
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	1.048	1.142	4.982	1.359	1.294	5.367
Estonia	obs. slope	-0.262	0.390	-0.261	-0.253	0.378	-0.237
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	2.944	2.881	6.428	0.915	0.691	4.607
Finland	obs. slope	-0.256	0.382	-0.259	-0.247	0.371	-0.235
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	1.468	1.475	5.240	0.770	0.918	5.352
Italy	obs. slope	-0.271	0.407	-0.283	-0.266	0.400	-0.262
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	3.132	3.842	15.117	3.022	3.928	3.682
Peru	obs. slope	-0.261	0.389	-0.262	-0.252	0.378	-0.237
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	2.609	2.766	6.390	0.509	0.555	4.729
Poland	obs. slope	-0.261	0.385	-0.255	-0.252	0.373	-0.231
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	2.483	1.852	3.005	0.460	0.521	6.284
Slovakia	obs. slope	-0.266	0.396	-0.261	-0.256	0.381	-0.234
	$\frac{\lvert\beta-\hat{k}\rvert}{se}$	4.407	4.526	6.479	1.622	1.486	7.077

I Third Paper: The Nemčok-Šedo Dataset

I am grateful to Dr Miroslav Nemčok for sharing with me a dataset of electoral system quantities for 560 elections in 40 countries that employ simple electoral systems, which expands on that used by Taagepera (2007) for testing the SPM. Table I.1 details the number of election and time range covered, as well as minima and maxima of M (mean district magnitude), S (assembly size), N_S (effective number of parties) and σ_1 (seat share of the largest party), grouped by country.

Table I.1: The Nemčok-Šedo dataset of inter-party quantities: descriptive statistics.

Country	# elections	Time Range	M (min-max)	S (min-max)	N_S (min-max)	σ_1 (min-max)
Armenia	2	2007–2012	3.2	131	2.7–3.4	0.49–0.53
Australia	27	1946–2013	1	74–150	2.2–3.2	0.39–0.6
Austria	20	1949–2013	4.3–20.3	165–183	2.1–4.6	0.28–0.52
Belgium	22	1946–2014	6.7–13.6	150–212	2.5–10.1	0.15–0.51
Bulgaria	9	1991–2017	7.7	240	2.4–5.1	0.34–0.57
Canada	23	1945–2015	1	245–338	1.5–3.2	0.4–0.78
Croatia	6	2000–2016	12.6–12.8	151–153	2.7–3.5	0.39–0.53
Cyprus	5	1996–2016	9.3	56	3.5–4.5	0.32–0.36
Czech Republic	8	1990–2013	14.3–25	200	2.3–5.6	0.25–0.62
Denmark	27	1945–2015	5.9–14.9	148–179	3.7–7.2	0.26–0.42
Estonia	7	1992–2015	8.4–9.2	101	3.8–5.9	0.28–0.41
Finland	20	1945–2015	12.5–15.4	200	4.6–5.8	0.22–0.32
France	17	1946–2012	1–4.4	475–618	1.8–6.2	0.2–0.75
Iceland	22	1946–2016	1.9–10.5	52–63	3.2–5.3	0.29–0.42
Ireland	20	1948–2016	3.4–4	144–166	2.4–4.6	0.32–0.57
Israel	20	1949–2015	120	120	3.1–8.7	0.22–0.47
Italy	12	1946–1992	17.4–19.7	556–630	2.6–5.7	0.33–0.53
Japan	12	1960–1993	3.9–4	467–512	2–4.1	0.44–0.64
Latvia	6	1998–2014	20	100	3.9–6	0.23–0.33
Luxembourg	16	1945–2013	12.8–16	51–64	2.7–4.3	0.31–0.5
Macedonia	8	1990–2016	1–20	120–123	1.9–6	0.32–0.72
Malta	6	1996–2017	5–5.3	65–69	2–2	0.51–0.57
Moldova	7	1994–2010	101–104	101–104	1.8–3.4	0.4–0.7
Montenegro	9	1990–2012	5.1–85	71–125	2.1–3.2	0.47–0.66
Netherlands	22	1946–2017	100–150	100–150	3.5–8.1	0.21–0.36
New Zealand	17	1946–1993	1	80–99	1.8–2.2	0.51–0.69
Norway	18	1945–2013	5.2–8.9	150–169	2.7–5.4	0.26–0.57
Poland	8	1991–2015	8.8–12.4	460	2.7–10.9	0.13–0.51
Portugal	15	1975–2015	10–11.9	230–263	2.2–4.2	0.35–0.59
Romania	5	1990–2004	7.9–9.7	332–396	2.2–4.8	0.34–0.66
Serbia	11	1990–2016	1–250	250	1.6–4.9	0.29–0.78
Slovakia	9	1990–2016	37.5–150	150	2.9–6.1	0.24–0.55
Slovenia	8	1990–2014	5.7–11.2	80–90	4.1–8.2	0.17–0.4
South Africa	5	1994–2014	44.4	400	2–2.3	0.62–0.7
Spain	13	1977–2016	6.7	350	2.3–4.1	0.35–0.58
Sweden	21	1948–2014	8.2–12.5	230–350	2.8–5	0.32–0.54
Switzerland	18	1947–2015	7.7–8	194–200	4.7–6.8	0.22–0.32
Turkey	18	1950–2015	4.2–9.1	400–610	1.2–4.9	0.25–0.92
Ukraine	3	1994–2007	1–450	450	3.1–3.4	0.19–0.41
United Kingdom	20	1945–2017	1	625–659	2–2.6	0.47–0.63
United States	18	1948–2016	1	435–437	1.8–2	0.51–0.68

J Third Paper: precision analysis for v_s predictions

As noted in the discussion of methodological choices(footnote 22 in section 5.5.2), the analysis presented in section 5.6.2 is restricted to v_1 and N_c insofar as these quantities have intuitive inter-party analogues in σ_1 and N_S. Diagnostics of deviation-from-prediction can however be conducted on the v_s list-level and district-level models as well. Consistently with the main analysis, the index d is computed as $\log_{10}\left(\frac{v_s}{(s^4cp)^{-1/4}}\right)$ for the list-level model and as $\log_{10}\left(\frac{\tilde{v}_s}{(prM^{\frac{5}{2}})^{-1/4}}\right)$ for the district-level model tested on district medians. Table J.1 presents median and mean values of absolute discrepancy, with the associated factors of error, and the share of observations that fall withing the 'tolerable error' band. Figure J.1 plots the values of the index d against the base products $(s^4cp)^{-\frac{1}{4}}$ and $(prM^{\frac{5}{2}})^{-\frac{1}{4}}$.

Figure J.1: Comparison of deviation from prediction of the district-level model for the means values of the dependent variables: $\bar{v}_1$, $\bar{N}_c$, and $\bar{v}_s$. Dashed lines represent values of d corresponding to values where the observed value is either twice or half the prediction.

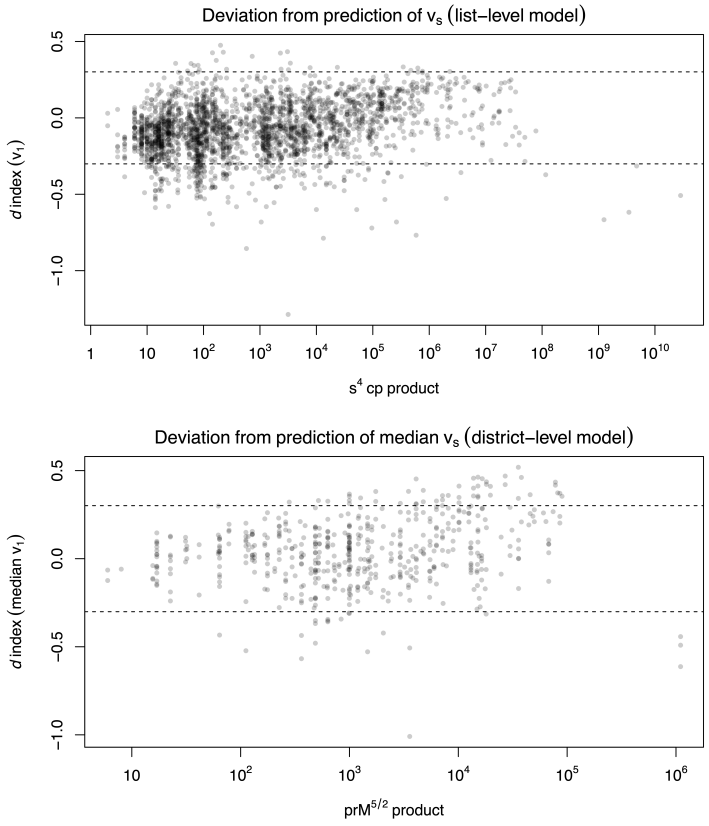
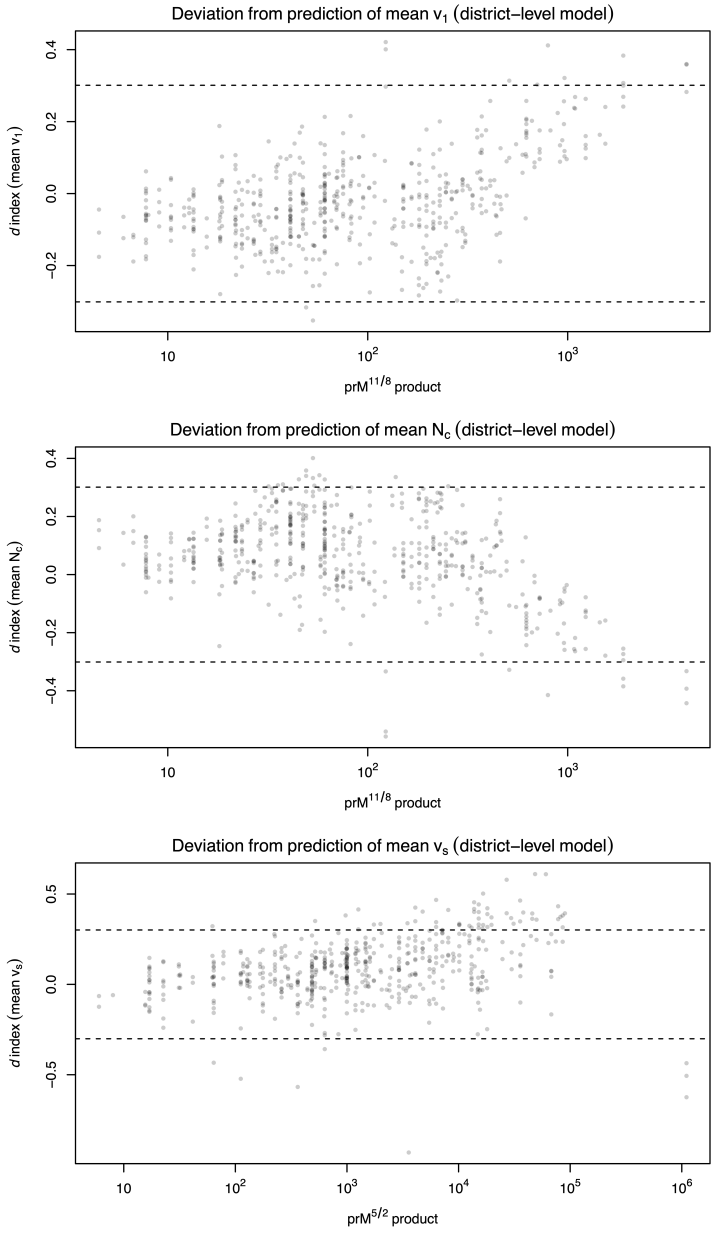


Table J.1: Summary indicators of deviation from prediction: models for v_s and $\tilde{v}_s$.

	median of $\|d\|$		mean of $\|d\|$		share $d < log_{10}(2)$
	value $\|d\|$	% error	value $\|d\|$	% error	and $d > log_{10}(0.5)$
List-level model	0.122	32.5%	0.147	40.2%	90.6%
District-level model	0.121	32.2%	0.152	42%	87.3%

K Third Paper: precision analysis for district means

Figure K.1: Comparison of deviation from prediction of the district-level model for the means values of the dependent variables: $\bar{v}_1$, $\bar{N}_c$, and $\bar{v}_s$. Dashed lines represent values of d corresponding to values where the observed value is either twice or half the prediction.



Precision diagnostics for the district-level model may also be computed on district-level means of v_1, N_c and v_s, as opposed to the median values presented in the main analysis and in section J of the Appendix. Figure K.1 plots the distribution of the values of d against the base products $prM^{\frac{11}{8}}$ and $prM^{\frac{5}{2}}$. Diagnostics analogous to those presented in the main analysis are presented in table K.1.

Table K.1: Summary indicators of deviation from prediction: models for $\bar{v}_1$, $\bar{N}_c$ and $\bar{v}_s$.

	median of $\|d\|$		mean of $\|d\|$		share $d < log_{10}(2)$
	value $\|d\|$	% error	value $\|d\|$	% error	and $d > log_{10}(0.5)$
$\tilde{v}_1$	0.081	20.4%	0.098	25.3%	97.8%
$\tilde{N}_c$	0.109	28.7%	0.126	33.6%	95.8%
$\tilde{v}_s$	0.113	29.6%	0.142	38.8%	88.6%