A Dynamic Duverger's Law

Jean Guillaume Forand* and Vikram Maheshri†

October 10, 2013

Abstract

Electoral systems promote strategic voting and affect party systems. Duverger (1951) proposed that plurality rule leads to bi-partyism and proportional representation leads to multi-partyism. We show that in a dynamic setting, these static effects also lead to a higher option value for existing minor parties under plurality rule, so their incentive to exit the party system is mitigated by their future benefits from continued participation. The predictions of our model are consistent with multiple cross-sectional predictions on the comparative number of parties under plurality rule and proportional representation. In particular, there could be more parties under plurality rule than under proportional representation at any point in time. However, our model makes a unique time-series prediction: the number of parties under plurality rule should be less variable than under proportional representation. We provide extensive empirical evidence in support of these results.

1 Introduction

The relationship between electoral systems and the number of parties contesting elections has become a classic topic of study in political science. Duverger (1951), who first formulated the question precisely, postulated the 'law' that plurality rule leads to two-party competition and the complementary 'hypothesis' that plurality rule with a runoff and proportional representation favor multi-partism (see Benoit (2006) and Riker (1982)). Empirically, party systems are not observed to be particularly stable over time: irrespective of electoral system, there is substantial longitudinal variation in the number of parties active in a country. However, existing research, both theoretical (e.g., Feddersen (1992), Palfrey (1989)) and empirical (e.g., Lijphart (1994), Taagepera and Shugart (1989)), has focused overwhelmingly on static, cross-sectional environments.

*Department of Economics, University of Waterloo.

†Department of Economics, University of Houston. We would like to thank Scott Legree for excellent research assistance.

Duverger supported his law by appealing to a dynamic process in which the number of parties is winnowed down by the combined impact of mechanical and psychological effects: plurality rule systematically underrepresents minor parties (mechanical), and in anticipation of this fact, strategic voters gradually desert all but two parties (psychological). In response, Chhibber and Kollman (1998) argued that when "accounting for changes in the number of national parties over time within individual countries, however, explanations based solely on electoral systems [...] are strained. These features rarely change much within countries, and certainly not as often as party systems undergo change in some countries." As important features of political environments (voters' preferences, salient issues, party leaderships) evolve over time, changes in the number of parties over time should be expected. This remark still leaves open the possibility that different electoral systems induce systematically different party system dynamics.

In this paper, we theoretically and empirically analyze the dynamic implications of the electoral incentives underlying Duverger's Law. We make two primary contributions. First, we develop a simple dynamic model of partisan politics that implies that plurality rule elections generate lower variability in the number of national parties over time (or *partisan churn*) than more proportional systems. Second, we offer empirical support for this hypothesis by analyzing the relationship between partisan churn and the disproportionality of electoral systems in a panel of 54 democracies since 1945. As predicted by the model, we find that more disproportional electoral systems are robustly associated with less entry of new parties and less exit of old parties. Together, this constitutes support for a reinterpretation of Duverger's vision of party dynamics. Notably, our model does not make unambiguous static predictions of the relationship between the number of competing parties and the disproportionality of electoral systems. This point has been previously made in static theoretical models (Morelli (2004)) and reconciles empirical findings presented here and elsewhere.

In our model, parties function as vehicles to promote the preferred policies of ideologically motivated activists. Parties are formed, maintained, and possibly disbanded by their activists. Supporting a party is costly as it requires the resources necessary to run a serious campaign: recruiting good candidates, mobilizing party volunteers and raising advertising funds. In view of the critique of Chhibber and Kollman (1998), the key dynamic ingredient of the model is a stochastic political environment: for any number of reasons, the support garnered among the voters by the various policies preferred by the activists may evolve over time. It follows that activists' incentives to support parties to represent them may also evolve, so activists whose policy goals are currently out of favor with voters may disband an existing party in the hopes of forming a new party in the future when voters become more receptive. Since forming a new party is costlier than maintaining an existing party, being currently represented by a party generates an option value to the activists who support it.

We take the perspective that party systems adapt to changing political circumstances by allowing the formation of parties that champion policy positions that were not represented in previous elections. In the model, the evolution of the political environment drives the entry of new parties and exit of existing parties. In practice, variability in party systems often involves transformations of existing parties with a corresponding reshuffling of their leadership and base. Party exit rarely takes the form of an outright dissolution but rather of a merger with ideologically compatible opponents. For example, the Liberal Democrats in the United Kingdom were formed in 1988 through the combination of the Liberal and Social Democratic parties, and the Christian Democratic Appeal was formed in the Netherlands in 1977 through the merger of three mainstream Christian parties. Similarly, the creation of a new party often occurs through the splintering of an existing party (e.g., the Left parties in Germany in 2007 and in France in 2008 combined various elements from existing parties on the left of the political spectrum).

Under plurality rule, the static mechanical and psychological effects favor the exit of parties with low current (anticipated) voter support and inhibit the entry of new parties. While there is some debate on whether these effects can be separately identified (see Benoit (2002)), the importance of their combined effect has been extensively documented. At the country level, the effective number of parties either contesting elections or represented in legislatures is positively associated to various measures of the proportionality of electoral systems such as average district magnitude (see Blais and Carty (1991), Lijphart (1994), Neto and Cox (1997), Ordeshook and Shvetsova (1994) and Taagepera and Shugart (1989)).[1] At the electoral district level, measuring the importance of strategic voting in an electoral district of magnitude M typically involves comparing the votes obtained by the candidates with the $M+1$ and $M+2$-ranked number of votes, i.e., the election's runner-up and second runner-up (see Cox (1997) and Fujiwara (2011)). As noted by Cox (1997), in an equilibrium in which a district's voters coordinate onto at most one non-winning alternative, the ratio of votes for candidates with ranks $M+2$ and $M+1$ should be zero. Interestingly, Cox (1997) finds evidence that the proportion of districts with electoral outcomes approaching this 'Duverger' outcome shrinks as the district magnitude M increases, suggesting that the incentives promoting, and/or the effectiveness of, strategic voting is reduced under more proportional electoral systems. These empirical results motivate the two key assumptions that differentiate plurality rule elections from proportional elections in our model. First, under plurality rule, a party with a small expected vote share in the current election suffers a *minority penalty* to its realized vote share. This electoral disadvantage of small parties under plurality combines the impacts of the static mechanical and psychological effects. Second, given any expected vote share, a newly-formed party under plurality

[1] District magnitude is defined as the number of representatives elected in a district.

rule suffers an *entry penalty* to its realized vote share. The incentives for strategic voting, which are strongest under plurality rule, reflect voters' attempts to coordinate onto competitive candidates. Since past voting behavior is likely to facilitate coordination, we posit that barriers to entry faced by new parties are higher under plurality rule.

Our model's novel dynamic insights come from combining (i) the variation in maintenance costs for minor parties and entry costs for new parties across electoral systems with (ii) the party support decisions of forward-looking activists represented by currently unpopular parties. Under plurality rule, the current cost to minor parties of maintaining their position is high, which incentivizes party exit. On the other hand, the option value of their position, which reflects the higher future costs of forming new parties, is also high, which incentivizes party maintenance. Under proportional representation, activists can respond more flexibly to changes in their current political circumstances by disbanding the parties they support in unfavorable political environments and forming new parties when, for example, new issues become salient. As the previous discussion suggests, our model makes no prediction of the number of parties in a given country at a given point in time. In fact, under plurality rule, we derive equilibria in which, in all elections at any time, there are at least as many active parties as under proportional representation. However, all equilibria under plurality rule feature less longitudinal variation in the number of active parties than the unique equilibrium under proportional representation: irrespective of the current number of parties, partisan churn under plurality rule is lower than under proportional representation.

We provide empirical support for our model with an analysis of competitive party behavior over time in democracies with varying levels of electoral proportionality. Our data come from the Constituency-Level Elections (CLE) Dataset from which we construct an unbalanced panel of elections in 54 countries since 1945 (Brancati (accessed 2013)). Our key empirical finding is that the proportionality of a country's electoral system is robustly correlated to the level of partisan churn observed in its elections. Highly proportional electoral systems such as Israel and Belgium feature elections with systematically greater entry and exit of parties than highly disproportional electoral systems such as the United States and Mexico. We subject this finding to a number of robustness checks and find that the dynamic relationship persists. On the other hand, we do not find strong evidence in favor of the static prediction of Duverger's Law. That is, although we find a positive relationship between the proportionality of an electoral system and the number of parties that compete in a given election, this association is not statistically significant. Hence, while the oft-cited 'exceptions' to Duverger's Law (e.g., Austria, Canada) blur the cross-sectional link between electoral rules and the number of parties as predicted by a number of theoretical models including ours, their longitudinal relationship, which is the novel prediction of our dynamic model, is quite strong.

In the terminology of Shugart (2005), ours is a 'macro level' study in that we focus on parties' entry and exit decisions in elections to the national parliament. This aggregation is necessary, and our hypothesis cannot be evaluated at the electoral district-level: a serious party either participates in elections in a large number of districts or risks failing to be considered as a legitimate national party. In fact, Fujiwara (2011) demonstrates this when he finds that the electoral system (plurality versus plurality with a runoff) has no impact on the identities of the parties competing for the mayoralty of Brazilian cities. He attributes this to the fact that serious candidates are affiliated to a major national party, and all serious national parties field candidates in most mayoral elections. It has long been noted that the results of Duverger (1951) are naturally established at the district level, and that his arguments establishing the 'linkage' of electoral systems' effects on the number of parties at the district level with the number of parties on the national stage are incomplete (see Cox (1997)). While a growing number of empirical studies address this linkage problem (see Chhibber and Kollman (1998), Chhibber and Murali (2006), Cox (1997)), theoretical investigations of Duverger's results have mostly focused on a single electoral district. In an important exception, Morelli (2004) shows that Duverger's predictions can be reversed in a multi-district setting if there is enough heterogeneity across districts. Our model shows that even abstracting from the linkage problem and considering a single district, the cross-sectional predictions of Duverger can be reversed solely due to the dynamic incentives of parties' supporters. The key contribution in this paper is that we recover a unique time series prediction.

While Duverger (1951) couched his arguments in dynamic terms, intertemporal approaches to the study of comparative political systems are rare. Cox (1997) highlights the importance of the dynamic incentives of parties and politicians for understanding the limits to Duverger's predictions, but he does not propose a particular model. Fey (1997) studies a dynamic process involving opinion polls to show that non-Duverger equilibria of the standard static model are unstable. We are not aware of any other theoretical paper embedding the study of the number of parties in a dynamic framework. Some recent empirical studies have focused on the dynamics of the number of parties. Chhibber and Kollman (1998) show that in the United States and India, the number of parties decreased in periods in which the central government assumed a larger role. This result, which compares countries with plurality elections, is focused on providing conditions which support the linkage from district to the national level. Reed (2001) provides evidence that at the district level elections became increasingly bipartisan in Italy following a change of voting rule in 1993. However, Gaines (1999) finds little evidence of a trend towards local two-partism in a longitudinal analysis of Canadian elections (see also Diwakar (2007) for the case of India).

2 The Dynamics of Party Entry and Exit: Model

2.1 Setup

Elections are held over an infinite horizon. Following an election at time $t = 1, 2, ...$, the winning party selects a policy $x^t \in \{x_{-1}, x_0, x_1\}$, where $x_{-1} < x_0 < x_1$. A party j can be of one of three types in $\{-1, 0, 1\}$ (e.g., left, middle or right). Parties are formed and maintained by policy-motivated activists. Specifically, there are two long-lived activists of type -1 and 1, and in each period they simultaneously decide whether or not to support a party of their type to represent them. We make two simplifying assumptions that allow us to focus on the incentives of these two non-centrist activists to form, maintain and disband parties. First, we assume that parties are non-strategic: if in power, party j implements policy x_j. Second, we assume that a party of type 0 is present in all elections. This simple environment allows for rich dynamics for party entry and exit, as well as for party structures, which in any given election can feature one, two or three parties. The electoral rule, which we detail below, is either plurality rule or proportional.

At the beginning of each period, a *preference state* $s^t \in \{s_{-1}, s_0, s_1\}$ is randomly drawn. Preference states capture variability in the political environment, which generates an option value to parties that maintain their electoral presence and is absent from static models. We assume that preference states are identically and independently distributed across periods: let $Pr(s^t = s_0) = q$ and $Pr(s^t = s_1) = Pr(s^t = s_{-1}) = \frac{1-q}{2}$ for $q \in (0, 1)$.[2] Preference states have a straightforward interpretation: in state s_j, the party representing activist j is favored by voters. Specifically, define $\overline{p}$, p and $\underline{p}$ such that $1 \geq \overline{p} > p > \underline{p} \geq 0$ and $\overline{p} + p + \underline{p} = 1$, and let p_j^t represent the *share of voters in the population who support policy* x_j in period t, with $p_{-1}^t + p_0^t + p_1^t = 1$. For the two non-centrist policies $x_j \in \{x_{-1}, x_1\}$, we define

$$p_j^t = \begin{cases} \overline{p} & \text{if } s^t = s_j, \\ \underline{p} & \text{if } s^t = s_{-j}, \\ \frac{\overline{p}+\underline{p}}{2} & \text{if } s^t = s_0. \end{cases}$$

Note that this implies that when the voters have non-centrist preferences (i.e., $s^t \in \{s_{-1}, s_1\}$), the share of voters who support the centrist policy x_0 is p.

While fraction p_j^t denotes the expected popularity of policy x_j in the election at time t, this policy may not be championed by a party if the activist of type j does not support a party. Conversely,

[2] We could allow for persistence in electoral states, although this would add computational complexity without affecting our central conclusions. Likewise, the simplifying assumption that non-centrist preference states s_1 and s_{-1} occur with equal probability allows us to exploit symmetry, but it is not essential.

a party championing policy x_j may have an expected popularity that exceeds p_j^t since it may draw support from voters whose preferred policy is not championed by party at t. A *party structure* ϕ^t lists the non-centrist parties supported by their activists in the current election: formally, $\phi^t \in 2^{\{-1,1\}}$. If a party supported by a non-centrist activist $j \in \{-1, 1\}$ is active under ϕ^t, then its *electoral support*, P_j^t, is equal to p_j^t. If instead this activist fails to support a party at t, the centrist party 0 collects the support of voters which would have supported a party of type j. Specifically, we define the support of party 0 under ϕ^t as

$$P_0^t = p_0^t + p_{-1}^t \mathbb{I}_{-1 \notin P^t} + p_1^t \mathbb{I}_{1 \notin P^t},$$

where $\mathbb{I}$ is the indicator function

The legislative power of activists depends on their support among the voters and on whether or not they are represented in elections by a party, but it is also mediated by the electoral system. At the conceptual level, we can represent plurality and proportional electoral systems as leading to different mappings from the distribution of voter support for parties into the distribution of seats in the legislature and corresponding policy outcomes (see Faravelli and Sanchez-Pages (2012) and Herrera et al. (2012)). On average, legislative policy outcomes under proportional representation should be more representative of voters' views as expressed by vote shares, while policy outcomes under plurality rule are more heavily tilted towards the views of plurality voters. We model this mapping in a reduced form, adopting a probabilistic voting approach that maps support shares among voters for active parties into these parties' probabilities of winning the election and implementing their ideal policies, which we interpret as obtaining decisive power in the legislature. Although this presents an incomplete view of legislative policy-making, our goal is to construct a minimal dynamic model of elections that predicts the observed patterns in party entry and exit documented in Section 3.

Under proportional representation, we assume that the *probability of winning of any active party* j is its support share among the voters P_j^t. Under plurality rule, we assume that the higher incentives for strategic voting impose coordination costs on small existing parties as well as on new parties of all sizes. First, given a non-centrist party $j \in \{-1, 1\}$ that is active at t when the preference state is s_{-j}, then if party $-j$ is also active party j bears a *minority penalty* to its probability of winning indexed by $\alpha \geq 0$. As discussed in the Introduction, we view this cost imposed on minor parties under plurality as representing the net effect of the mechanical effect due to the electoral formula and the psychological effect due to strategic voting as highlighted by Duverger (1951). Second, in any preference state at t, if a non-centrist activist j forms a new party and party $-j$ is active in both the election at $t-1$ and t, then party j bears an *entry penalty* to its probability of winning indexed by $\beta \geq 0$. This is a dynamic effect of increased incentives for strategic voting under plurality:

whether or not a party was active in past elections can act as a natural coordination device for voters. Specifically, under plurality rule, fix time t and suppose that the party structure in the current election is such that $\phi^t = \{-1, 1\}$. Then the *probability of winning of a non-centrist party* j is

$$P_j^t + \alpha \left[\mathbb{I}_{s^t = s_j} - \mathbb{I}_{s^t = s_{-j}} \right] - \beta \mathbb{I}_{j \notin \phi^{t-1}} \mathbb{I}_{-j \in \phi^{t-1}}.$$

Meanwhile, if $\phi^t = \{j\}$, then the probability with which party j wins is P_j^t. To ensure that, for any s^t an active party j has a non-negative winning probability, we assume that $\alpha + \beta \le \underline{p}$. Note that our formulation assumes that any coordination costs imposed on party j benefit only party $-j$, which implies that in any preference state, the probability of winning of party 0 under plurality rule is P_0^t.

Activists are risk-neutral and have single-peaked preferences over feasible policies with a non-centrist activist of type j having ideal policy x_j. Given any non-centrist activist, let $\overline{u}$ be its stage payoff to its preferred policy, u be its stage payoff to its second-ranked policy, and $\underline{u}$ be its stage payoff to its third-ranked policy with $\overline{u} > u > \underline{u}$. Supporting parties is costly for activists, although forming a new party is costlier than maintaining an existing party. This wedge between the cost of maintaining an existing party and the cost of forming a new party generates an option value to existing parties for activists. Specifically, at time t, if $j \in \phi^{t-1}$, then the party maintenance cost to activist j in the electoral cycle at t is $\underline{c}$. If instead $j \notin \phi^{t-1}$, then no party represented activist j in the previous election and the party formation cost at t to activist j is $\overline{c} > \underline{c}$. Activists discount future payoffs by a common factor of δ, and make party support decisions to maximize their expected discounted sum of payoffs, which in any election consists of the expected difference between its benefits from the policy implemented by the winning party and its party formation costs (where the expectation is over electoral outcomes).

2.2 Strategies and Equilibrium

We focus on Markov perfect equilibria in pure strategies in which activists condition their party formation and maintenance decisions at time t on the payoff-relevant *state* (s^t, ϕ^{t-1}): the current preference state and the previous party structure. For a non-centrist activist j, a strategy is $\sigma_j : \{s_{-1}, s_0, s_1\} \times 2^{\{-1,1\}} \to \{0, 1\}$, where $\sigma_j(s, \phi) = 1$ indicates that the activist supports a party in preference state s given party structure ϕ inherited from past periods. Let $V_j(s, \phi; \sigma)$ denote the expected discounted sum of payoffs to activist j under profile $\sigma \equiv (\sigma_{-1}, \sigma_1)$ conditional on state

(s, ϕ). Profile σ^* is a *Markov perfect equilibrium* if, for all states (s, ϕ) and all profiles (σ_{-1}, σ_1),

$$V_{-1}(s, \phi; \sigma^*) \geq V_{-1}(s, \phi; (\sigma_{-1}, \sigma_1^*)) \text{ and}$$
$$V_1(s, \phi; \sigma^*) \geq V_1(s, \phi; (\sigma_{-1}^*, \sigma_1)).$$

From now on, the term equilibrium refers to Markov perfect equilibrium. Restricting attention to strategies in which activists condition only on payoff-relevant elements of histories of play limits the possibilities for intertemporal coordination between activists. In our model, as will be clear below, it also ensures that equilibrium behavior is relatively simple.

2.3 Results

The comparative equilibrium dynamics of party systems under both electoral systems depends critically on the values of party formation and maintenance costs $(\overline{c}, \underline{c})$, coordination costs (α, β), and policy payoffs $(\overline{u}, u, \underline{u})$. For example, if $\overline{c} > \overline{u}$, then under both electoral systems no non-centrist party ever forms in any equilibrium. Conversely, if $\underline{c} = 0$ and $\underline{p} > 0$, then under both electoral systems no existing non-centrist party is ever disbanded in any equilibrium. Our interest lies in those regions of the parameter space in which any equilibrium party maintenance by current minority activists is due solely to dynamic incentives. That is, we restrict attention to parameter values such that, in the static stage game with preference state s_{-j}, activists of type j prefer to disband their party when anticipating that a non-centrist party j will contest the election.

We first present our results for proportional representation. Our aim is to show that the lower coordination costs under proportional representation allow activists to better tailor their party formation and maintenance decisions to the current preference state by supporting parties when voters' preferences favor their policy positions and disbanding parties when they do not. To this end, we introduce a strategy profile in which non-centrist activists support parties if and only if the current electoral state does not favour the activist on the other side of the political spectrum. Specifically, define profile σ^{PR} such that, for any non-centrist activist j and party structure ϕ,

$$\sigma_j^{PR}(s, \phi) = \begin{cases} 1 & \text{if } s \in \{s_j, s_0\} \\ 0 & \text{if } s = s_{-j}. \end{cases}$$

In the following result, we identify conditions under which the strategy profile σ^{PR} is an equilibrium under proportional representation. Furthermore, we show that under these same conditions no other equilibrium exists.[3]

[3] All proofs are in Appendix A.

Proposition 1. *Suppose that*

$$\overline{c} < \frac{1-\overline{p}}{2}[\overline{u}-u], \tag{1}$$

and that

$$\underline{c} > \underline{p}[\overline{u}-u] + \delta\frac{1+q}{2}[\overline{c}-\underline{c}]. \tag{2}$$

Then σ^{PR} is the unique Markov perfect equilibrium under proportional representation.

Condition (1) ensures that a non-centrist activist j always supports a party in s_j and s_0, so that the only remaining question is whether or not the activist will support a party in s_{-j}. Note that under condition (2), we have that $\underline{p}[\overline{u}-u] - \underline{c} < 0$, so in the stage game with preference state s_{-j}, activist j prefers disbanding an existing party to maintaining it. However, maintaining an existing party in s_{-j} has an associated option value realized in s_j and s_0, which is derived from the cost savings for supporting a party in those states. Condition (2) ensures that under proportional representation, the immediate cost savings from disbanding an existing party dominates the option value of supporting it through an unfavourable election. Conditions (1) and (2) uniquely pin down the optimal party formation and maintenance decisions of both non-centrist activists, so that no other equilibrium can exist. Also, note that while the equilibrium σ^{PR} is in symmetric strategies, we impose no ex ante symmetry restriction on equilibria.

We now turn to our results under plurality rule. Our aim is to show that in those regions of the parameter space identified in Proposition 1, the coordination costs imposed on parties under plurality rule lead activists' party formation and maintenance decisions to display more persistence than under proportional representation. Accordingly, we focus attention on strategy profiles in which activists support *existing* parties if and only if the preference state does not favour the activist on the other side of the political spectrum. Contrary to the case of profile σ^{PR} under proportional representation, entry penalties induce activists to form *new* parties only when the preference state favours them. Specifically, we restrict attention to profiles σ^{PL} with the property that for all non-centrist activists j,

$$\sigma_j^{PL}(s,\phi) = \begin{cases} 1 & \text{if } s = s_j, \text{ or if } s = s_0 \text{ and } \phi \neq \{-j\}. \\ 0 & \text{if } s \in \{s_0, s_{-j}\} \text{ and } \phi = \{-j\}. \end{cases} \tag{3}$$

The key question is whether activist j supports an existing party when the preference state favours its opponent. On the one hand, minority penalties increase the cost of maintaining a party in unfavourable electoral circumstances. On the other hand, entry penalties increase the option value of a party that is maintained even through a string of lost elections. We consider two alternatives.

Profile $\overline{\sigma}^{PL}$ denotes the strategy profile respecting (3) with maximal participation:

$$\overline{\sigma}_j^{PL}(s,\phi) = 1 \text{ if } s = s_{-j} \text{ and } j \in \phi,$$

while profile $\underline{\sigma}^{PL}$ denotes the strategy profile respecting (3) with minimal participation:

$$\underline{\sigma}_j^{PL}(s,\phi) = 0 \text{ if } s = s_{-j} \text{ and } j \in \phi.$$

In the following result, we identify conditions under which $\overline{\sigma}^{PL}$ and $\underline{\sigma}^{PL}$ are equilibria under plurality rule.[4] These conditions will depend on the entry penalty β being bounded above and below. These upper and lower bounds, denoted $\overline{\beta}$ and $\underline{\beta}$ respectively, are functions of all the parameters of the problem except the minority penalty α, and they are derived in Appendix A.

Proposition 2. *Suppose that* (1) *and* (2) *hold and that* $\beta \in (\underline{\beta}, \overline{\beta})$. *Then there exist* $\underline{\alpha}, \overline{\alpha} \in [0, \underline{p} - \beta]$ *such that* $\underline{\sigma}$ *is a Markov perfect equilibrium whenever* $\alpha > \underline{\alpha}$ *and* $\overline{\sigma}$ *is a Markov perfect equilibrium whenever* $\alpha < \overline{\alpha}$. *Furthermore,* $\underline{\alpha} \geq \overline{\alpha}$.

Our dynamic model provides no robust cross-sectional predictions on the number of parties under different electoral systems. In any given election under proportional representation, there could be either two or three parties competing (under σ^{PR}). Under plurality, our model allows for the standard Duverger prediction of a two-party system (under $\underline{\sigma}^{PL}$), although the identities of the parties change over time as voters' preferences evolve, but it also allows for a non-Duverger equilibrium in which three parties are always present (under $\overline{\sigma}^{PL}$). However, our model does provide a robust dynamic prediction: there is greater variation in the number of active parties in equilibrium σ^{PR} under proportional representation than under either of the equilibria $\underline{\sigma}^{PL}$ and $\overline{\sigma}^{PL}$ that we identify under plurality. To see this, first note that there is no variation in the number of parties under $\overline{\sigma}^{PL}$ as three parties contest all elections. To compare σ^{PR} and $\underline{\sigma}^{PL}$, note that under both equilibria, a transition from s_j to s_{-j} leads to the exit of the party representing activist j and the entry of the party representing activist $-j$. However, for other transitions in preference states, σ^{PR} generates more variability in the number of parties. A transition from s_j to s_0 leads to the entry of party $-j$ under σ^{PR} but not under $\underline{\sigma}^{PL}$, while a transition from s_0 to s_j always leads to the exit of party $-j$ under σ^{PR}. This party need not be active in this state under $\underline{\sigma}^{PL}$, in which case no exit can occur.

Although preference states are drawn independently across periods, party structures under plurality are history-dependent while party structures under proportional representation are not. Under

[4] Activist j's actions are not yet specified only if the preference state is s_{-j} and no activists supported parties in the previous elections (i.e., $\phi^t = \emptyset$). These histories only occur off the equilibrium path, and the details are in Appendix A.

σ^{PR}, the probability that a party representing activist j contests any election is $\frac{1+q}{2}$ (the probability that the preference state is either s_j or s_0) which does not depend on the realization of past preference states or party structures. Under plurality, party structures are fully persistent in the equilibrium $\overline{\sigma}^{PL}$, as no party ever exits. In the equilibrium $\underline{\sigma}^{PR}$, the probability that a party representing activist j contests an election at time t depends on whether or not this party contested an election at time $t-1$. Specifically, if $j \in \phi^{t-1}$, then party j contests the election at time t with probability $\frac{1+q}{2}$, the probability that the preference state is either s_j or s_0. On the other hand, if $j \notin \phi^{t-1}$, then it contests the election with probability $\frac{1-q}{2}$, the probability that the preference state transitions to s_j.

To understand the conditions under which $\underline{\sigma}^{PL}$, or alternatively $\overline{\sigma}^{PL}$, are equilibria, consider activist j in state $(s_{-j}, \{j\})$. Under $\underline{\sigma}^{PL}$, activist j disbands its current party and waits until the preference state returns to s_j before forming a new party to represent it. However, since in that case activist $-j$ will disband the party it forms in state $(s_{-j}, \{j\})$, activist j faces no entry penalty when it forms a new party. Hence, $\underline{\sigma}^{PL}$ provides incentives for activist j to disband its party in s_{-j} only if minority penalty α is sufficiently high to deter party maintenance. On the other hand, under $\overline{\sigma}^{PL}$ activist j supports its party and bears the minority penalty, which cannot be too high in order to provide incentive for party maintenance. For a given minority penalty α, the two profiles cannot both be equilibria. The lower bound $\underline{\beta}$ on the entry penalty ensures that these costs are high enough to prevent activists that are not represented by a party in centrist state s_0 from forming a new party. Note that such histories occur on the equilibrium path only under $\underline{\sigma}^{PL}$. The upper bound $\overline{\beta}$ on the entry penalty ensures that these costs are low enough that, under $\overline{\sigma}^{PL}$, non-centrist activist j is willing to form a new party in preference state s_j, in those histories off the equilibrium path in which this activist is not represented by a party. Note that for such histories under $\underline{\sigma}^{PL}$, activist j never bears entry penalties since no party representing activist $-j$ ever contests elections in preference state s_j.

Condition (2) does not play a role in the proof of Proposition 2. We include it in order to establish that the equilibria $\underline{\sigma}^{PL}$ and $\overline{\sigma}^{PL}$ can exist under plurality under parametric restrictions that ensure that σ^{PR} is the unique equilibrium under proportional representation. A simple example is sufficient to show that the conditions of Proposition 2 can be met. Suppose that $\delta \approx 1$, $(\underline{p}, p, \overline{p}) = (\frac{2}{10}, \frac{3}{10}, \frac{5}{10})$, $\beta = \frac{1}{10}$, $\overline{u} - u = 1$, $u - \underline{u} = \frac{3}{2}$, $\overline{c} = \frac{3}{8}$ and that $\underline{c} = \frac{5}{16}$. Given these parameters, it can be computed that all the conditions in Proposition 2 are respected, and that furthermore $\underline{\alpha} < \underline{p} - \beta$ and $\overline{\alpha} > 0$, so that both $\underline{\sigma}^{PL}$ and $\overline{\sigma}^{PL}$ can be equilibria for that value of β, depending on the value of α.

3 The Dynamics of Party Entry and Exit: Emprical Findings

The key empirical implications of our model concern the dynamic relationship between electoral systems and partisan competition. In particular, our model predicts that more disproportional electoral systems should experience less churn as parties are less likely to enter and exit elections in these systems. To test these predictions, we use the Constituency-Level Elections (CLE) Dataset (Brancati (accessed 2013)), which contains information on the vote shares and seat shares of all political parties that participated in a broad sample of national democratic elections. Our empirical analysis consists of estimating the relationship between the disproportionality of an electoral system and the dynamics of its party system through party entries and exits. We do not ascribe a causal interpretation to any portion of our empirical anlysis as our aim is simply to provide robust evidence that is consistent with the central predictions of our model.

There are two main measurement issues that we must address in order to conduct our anlaysis. First, we need a concise measure of the disproportionality of an electoral system, which is determined by institutional characteristics such as electoral laws in a potentially complex manner. Second, we need an appropriate measure of party entry and exit. A key difficulty here is that electoral systems differ in their number of districts with more proportional systems having less districts on average than plurality systems, and parties may be active in some districts and not others. This can be the case if, for instance, a party's support is regional in nature. Alternatively, a successful entry in a few districts may be a launching pad for a new national party.

To address the first issue, we use the least squares index of Gallagher (1991) to measure the disproportionality of an electoral system. This index, which has been widely used in empirical analyses of electoral systems, is a measure of the difference between parties' vote and seat shares in a given election.[5] In perfectly proportional electoral systems, parties' seat shares should be identical to their vote shares, while in less proportional systems front-running parties typically have seat shares exceeding their vote shares and lagging parties have seat shares well below their share of the votes. Formally, for a given election e in a given country c with J total parties, let p_{jce} be the vote share that party j receives, and let s_{jce} be the seat share that party j wins in the legislature. Then the disproportionality index for this election is given by

$$g_{ce} = \sqrt{\frac{1}{2}\sum_{j=1}^{J}(p_{jce} - s_{jce})^2} \tag{4}$$

where g_{ce} is an index that ranges from 0 to 1 with increasing values corresponding to more dispro-

[5]See also Lijphart (1994) and Taagepera and Grofman (2003).

portional elections. Because disproportionality is a property of the electoral institutions of country, it should should not vary either by electoral district or by election. Hence, we aggregate district electoral outcomes and compute the disproportionality index at the national level. Furthermore, we average the disproportionality index over all elections for a different country, i.e.,

$$G_c = \frac{1}{E_c} \sum_{e=1}^{E_c} g_{ce} \tag{5}$$

where E_c is the total number of elections that we observe for country c. Hence, G_c is our measure of the disproportionality of the electoral system of country c.

To address the issue of measuring party entry and exit, we proceed as follows. For any election e in country c, we denote the number of electoral districts D_{ce}, where district d contributes a fraction σ_{dce} of the total seats in the national legislature. A party is said to have *entered* in district d in election e if its vote share in that district in $e-1$ was less than 0.05 and its vote share in that district in e was greater than 0.05. Party exit is defined similarly.[6] Let n_{dce} and x_{dce} represent, respectively, the total number of entering and exiting parties in district d during election e in country c. The total number of entries N_{ce} in a given election is obtained by summing over all districts as

$$N_{ce} = \sum_{d=1}^{D_{ce}} n_{dce} \cdot \sigma_{dce}, \tag{6}$$

and the total number of exits X_{ce} can be defined similarly as

$$X_{ce} = \sum_{d=1}^{D_{ce}} x_{dce} \cdot \sigma_{dce}. \tag{7}$$

We weigh the number of entries in each district by that district's size in order to correct for the variability in the number of electoral districts across electoral systems. For example, Israel, which is considered to have an electoral system that is almost perfectly proportional, has a single electoral district, so one entry is recorded if a new party collects a share of 0.05 of votes at the national level. The United Kingdom, on the other hand, has all legislators elected by plurality rule in over six hundred electoral districts, so that one entry is recorded if a new party collects a share 0.05 of votes in every district. The emergence of a regional party that collects the threshold share of votes in, say, half of the country's districts, would be recorded as half an entry. In the absence of weighing district-level party entries and exits, the variability in party structures in plurality rule

[6]As a robustness check, we replicated our analysis replacing the 0.05 threshold for entry and exit with 0.01, 0.02 and 0.10 and obtained similar results.

systems would be dramatically overstated. Finally, the total net party movements in an election (i.e., the total amount of partisan churn), M_{ce}, is simply defined as the sum of entries and exits as

$$M_{ce} = N_{ce} + X_{ce}.$$

We construct these variables from the CLE, which contains detailed information on the identities of all parties that participated in a large number of elections in many countries since 1945.[7] In particular, the CLE documents the number of votes that each party received in each district of a given election and the number of legislative seats that they were awarded. With this information, it is straightforward to construct the measures described above. In Table 1, we present the countries in our sample grouped by disproportionality and the total number of elections we observe in each group.[8] Because all countries do not hold elections at the same frequency (and several countries were formed or ceased to exist since 1945) our data set constitutes an unbalanced panel.

In Table 2, we present a traditional, static test of Duverger's Law and explore the relationship between the disproportionality of electoral systems and the number of parties that compete in elections.[9] In the first specification, we find this relationship to be negative, which is consistent with our model (thought not dispositive), although this correlation is not statistically significantly different than zero.[10] This noisy evidence in favor of the static version of Duverger's Law is not considerably strengthened when we add additional control variables: in specifications 2 through 4, the basic finding of a negative but statistically insignificant relationship persists.[11]

In Table 3, we present our main empirical results, which explore the *dynamic* relationship between the disproportionality of electoral systems and the number of parties that compete in elections. In each set of four columns, we specify total entries, exits and net movements as the dependent variable respectively and the average Gallagher Index as the primary independent variable. The coefficient of interest on this variable is predicted to be negative by our model. For each dependent variable, we estimate four regressions, each of which includes different sets of control variables. In

[7]The CLE unfortunately does not contain data on all democratic elections since 1945. Indeed, no single source does. We use only those elections contained in the CLE for our analysis and do not supplement our dataset with data from other sources in order to maintin consistent reporting. We replicated our analysis using a similar (thought not identical) sample of elections from the Constituence Level Elections Archive (CLEA) data set and obtained similar results. We report results using the CLE because this is the dataset that has been primarily used to construct disproportionality indices (Gallagher and Mitchell (2005)).

[8]All tables are in Appendix B.

[9]The number of competing parties is computed in a similar manner to the numbers of entries and exits. That is, any party that receives a vote share over 0.05 in any election is counted. Our results are robust to alternative thresholds of 0.01, 0.02 and 0.10.

[10]This specification essentially reproduces the correlations in Table 3.4 of Lijphart (1994).

[11]Note that in specification 4, which contains the most control variables, the coefficient on disproportionality is significant at the 90% level.

all regressions, we specify all continuous variables in logarithms. By doing so, our parameter estimates are scale invariant. This ensures that electoral systems with many parties (which tend to be more proportional, per the static results) do not simply exhibit a large amount of partisan churn by construction. Rather, any such relationship between disproportionality and partisan churn should be interpreted as independent of the total number of parties.[12] Because elections may feature zero entries or exits, we transform all continuous variables x as $\log(1+x)$ in order to conserve data. Because the disproportionality index does not vary within country by construction, we cluster our standard errors at the country level to account for this induced multicollinearity.

In the first specification, we include no control variables. Consistent with our model, we estimate negative relationships between disproportionality and all three dependent variables, although only the relationship between disproportionality and total entries is statistically significant. In the second specification, we include dummies for each decade in order to absorb slowly varying global determinants of partisan political activity.[13] Again, we obtain negative estimates of our coefficients of interest, and these are statistically significant at least at the 95% level. In the third specification, we include regional dummies for European countries, African countries, and former republics of the USSR in order to absorb any regional determinants of political activity. The estimates of our coefficients of interest change very little and remain highly statistically significant. Finally, in the fourth specification, we flexibly control for the number of districts by including sixth order polynomials in D_{ce} and $\log D_{ce}$.[14] Because D_{ce} explicitly enters into our computation of entries and exits in equations (6) and (7), conditioning our regressions on the number of districts ensures that the coefficients of interest that we estimate are not simply mechanically determined by variation in this D_{ce}. Indeed, our coefficient estimates in specification 4 are negative, larger (in absolute value) and highly statistically significant at the 99% level.

In Table 4, we explore the extent to which our strongly consistent finding of a negative relationship between partisan churn and disproportionality relies on some of the empirical assumptions underlying our analysis. We first test the extent to which our results are sensitive to the construction of the country level disproportionality index, G_c, as given in equation (5). Instead of averaging the election level Gallagher indices by country, we instead construct an alternative country level

[12]We reestimate all regressions with flexible controls for the number of parties (sixth order polynomials in the number of parties and log(number of parties)) and obtain similar results to those presented, though they tend to be statistically significant only at the 90% level.

[13]Our decade dummies are defined for the periods 1940-49, 1950-59, ... , 2000-2009. We replicated our analysis defining decade dummies for all possible periods (e.g., 1948-1957, ...) and obtained results that were statistically indistinguishable from those presented.

[14]As a robustness check, we replicated our analysis with polynomials of all orders up to 10 in D_{ce} and $\log D_{ce}$ and obtained qualitatively similar negative and statistically significant estimates of our coefficients of interest at the 95% level.

disproportionality index $\hat{G}_c$ by using the disproportionality index of the first election in the sample for each country, i.e., $\hat{G}_c \equiv g_{c0}$, where $e = 0$ corresponds to the first election in the sample for country c. In the first specification, we use this alternative measure and find statistically significant results that are also statistically indistinguishable from the estimates obtained using G_c.

The fact that both the dependent variables and the disproportionality index are derived in part from electoral outcomes raises the concern that our results may reflect a spurious correlation. This concern is mitigated in part because we aggregate these variables according equations (5), (6) and (7). Nevertheless, we re-estimate our main regression in the second specification with an alternative proxy for disproportionality that is unrelated to electoral outcomes. Following Taagepera and Shugart (1989), we specify the proportionality of an electoral system as the average district magnitude, i.e., the total number of seats divided by the total number of districts. This variable is purely determined by electoral institutions, and hence should not suffer from this form of spurious correlation. More proportional systems will have districts of higher magnitude on average, so the results in the second specification are strongly consistent with our main dynamic results.

We also test the extent to which our results are sensitive to the logarithmic specification of continuous variables. In the third specification, we reestimate the regression with all variables specified linearly. As before, we obtain negative coefficeint estimates that are statistically significant at the 99% level. Although we hesitate to interpret these results because they might simply be driven by the fact that disproportional electoral systems tend to have more parties (and hence more churn as measured in levels), they are strongly robust to the inclusion of flexible controls for the total number of parties.

Finally, we test the extent to which our construction of the numbers of entries and exits in national elections are sensitive to the weighting described in equations (6) and (7). We compute unweighted aggregate district level entries and exits by effectively fixing $\sigma_{dce} = 1$. In countries with a single electoral district, such as Israel, this results in no change. However, in countries with many districts, this construction places greater weight on the entries and exits of smaller, regional parties. In the fourth specification, we find that the use of this alternative measure of entries and exits still yields negative estimates of the coefficients of interest that are statistically significant at the 99% level.[15] In sum, we take these results to constitute strong, consistent and robust evidence in favor of the dynamic predictions of our model.

[15] We note that when we estimate the regressions with unweighted measures of entries and exits but do not control flexibly for the number of districts in a country, we obtain very different estimates. This is consistent with the fact that the overcounting of entries and exits in the unweighted measures is exacerbated in countries with many districts. Hence, conditioning our estimates on the number of districts serves the additional purpose of mitigating this source of bias.

4 Conclusion

This paper presents a novel dynamic reinterpretation of Duverger's Law. We construct a minimal but transparent dynamic model that establishes that (i) static Duverger predictions on the comparative number of parties under plurality rule and proportional representation can be reversed when intertemporal incentives are taken into account and that (ii) a unique dynamic prediction can be recovered if we focus our attention on the comparative variation in the number of parties over time across electoral systems. We finds robust empirical support in favor of the latter prediction.

Since party formation and maintenance decisions are typically made on a national level, the dynamic predictions of our model can only be verified appropriately with cross-country elections data. Further, since electoral systems rarely change within countries, this hinders any attempt to attribute a causal effect of electoral systems on the evolution of the number of national parties. We consider the time-series correlations uncovered in this paper sufficiently novel, interesting and robust that the lack of a causal interpretation does not present a critical concern. However, we make a broader contribution in that we point to the interest of studying the comparative intertemporal properties of electoral systems. In future work, related questions along these lines may be amenable to causal inference as, for example, the study of the comparative importance of strategic voting in Fujiwara (2011) allowed causal claims about political forces leading to the cross-country predictions on the number of parties.

References

Benoit, K., 2002. The endogeneity problem in electoral studies: a critical re-examination of duverger's mechanical effect. Electoral Studies 21 (1), 35–46.

Benoit, K., 2006. Duverger's law and the study of electoral systems. French Politics 4 (1), 69–83.

Blais, A., Carty, R. K., 1991. The psychological impact of electoral laws: measuring duverger's elusive factor. British Journal of Political Science, 79–93.

Brancati, D., accessed 2013. Global Elections Dataset. New York: Global Elections Database, http://www.globalelectionsdatabase.com.

Chhibber, P., Kollman, K., 1998. Party aggregation and the number of parties in india and the united states. American Political Science Review, 329–342.

Chhibber, P., Murali, G., 2006. Duvergerian dynamics in the indian states federalism and the number of parties in the state assembly elections. Party Politics 12 (1), 5–34.

Cox, G. W., 1997. Making votes count: strategic coordination in the world's electoral systems. Vol. 7. Cambridge Univ Press.

Diwakar, R., 2007. Duverger's law and the size of the indian party system. Party Politics 13 (5), 539–561.

Duverger, M., 1951. Les partis politiques. Armand Colin.

Faravelli, M., Sanchez-Pages, S., 2012. (don't) make my vote count.

Feddersen, T. J., 1992. A voting model implying duverger's law and positive turnout. American journal of political science, 938–962.

Fey, M., 1997. Stability and coordination in duverger's law: A formal model of preelection polls and strategic voting. American Political Science Review, 135–147.

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

Gaines, B. J., 1999. Duverger's law and the meaning of canadian exceptionalism. Comparative Political Studies 32 (7), 835–861.

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

Gallagher, M., Mitchell, P., 2005. The politics of electoral systems. Cambridge Univ Press.

Herrera, H., Morelli, M., Palfrey, T. R., 2012. Turnout and power sharing.

Lijphart, A., 1994. Electoral systems and party systems: A study of twenty-seven democracies, 1945-1990. Oxford University Press.

Morelli, M., 2004. Party formation and policy outcomes under different electoral systems. The Review of Economic Studies 71 (3), 829–853.

Neto, O. A., Cox, G. W., 1997. Electoral institutions, cleavage structures, and the number of parties. American Journal of Political Science, 149–174.

Ordeshook, P. C., Shvetsova, O. V., 1994. Ethnic heterogeneity, district magnitude, and the number of parties. American journal of political science, 100–123.

Palfrey, T., 1989. A mathematical proof of duverger's law. In: Ordeshook, P. C. (Ed.), Models of strategic choice in politics. University of Michigan Press.

Reed, S. R., 2001. Duverger's law is working in italy. Comparative Political Studies 34 (3), 312–327.

Riker, W. H., 1982. The two-party system and duverger's law: An essay on the history of political science. The American Political Science Review, 753–766.

Shugart, M. S., 2005. Comparative electoral systems research: the maturation of a field and new challenges ahead. In: Gallagher, M., Mitchell, P. (Eds.), The politics of electoral systems. Cambridge Univiversity Press.

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

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

A Appendix: Proofs

Proof of Proposition 1. Note that (1) implies that under proportional representation, forming (or maintaining, since $\overline{c} > \underline{c}$) a party is uniquely stage optimal in preference state s_0 for party j, irrespective of whether activist $-j$ is represented by a party. Also, since $\overline{p} > \frac{1}{3}$, (1) implies that $\overline{c} \leq \overline{p}[\overline{u} - u]$, so that forming (or maintaining, since $\overline{c} > \underline{c}$) a party is uniquely stage optimal in preference state s_j for party j, irrespective of whether activist $-j$ is represented by a party. Finally, since $\overline{c} > \underline{c}$, it follows that, for any state (s, ϕ) and any equilibrium σ^*, $V_j(s, \phi \cup \{j\}; \sigma^*) \geq V_j(s, \phi; \sigma^*)$. Hence, in any equilibrium under proportional representation, it must be that $\sigma_j^*(s, \phi) = 1$ for all states such that $s \in \{s_0, s_j\}$.

It remains only to determine activists' equilibrium actions in preference state s_{-j}. Fix an equilibrium σ^* and consider a state (s_{-j}, ϕ) such that $j \in \phi$. If activist j disbands its party, its payoff is

$$V_j(s_{-j}, \phi; \sigma^*) = (1 - \overline{p})u + \overline{p}\underline{u} + \delta \mathbb{E} V_j(s', \{-j\}; \sigma^*)$$

If instead activist j maintains its party, let $V^d(s_{-j}, \phi; \sigma^*)$ be its payoff. We have that

$$V_j^d(s_{-j}, \phi; \sigma^*) = \underline{p}\overline{u} + pu + \overline{p}\underline{u} - \underline{c} + \delta \mathbb{E} V_j(s', \{-j, j\}; \sigma^*).$$

By our results from above, we have that, for any $s \in \{s_0, s_j\}$,

$$V_j(s, \{-j\}; \sigma^*) = V_j(s, \{-j, j\}; \sigma^*) - [\overline{c} - \underline{c}],$$

so that $V_j(s_{-j}, \phi; \sigma^*) > V_j^d(s_{-j}, \phi; \sigma^*)$ if and only if (2) holds. Note that (2) also implies that in state (s_{-j}, ϕ) such that $j \notin \phi$, activist j strictly prefers not to form a party. Hence, for any equilibrium σ^* under proportional representation, we have that $\sigma^* = \sigma^{PR}$. □

Proof of Proposition 2. Define $\underline{\beta}$ and $\overline{\beta}$ such that

$$\underline{\beta}[\overline{u} - \underline{u}] \equiv \frac{1}{1-\delta q}\left[\frac{1-\overline{p}}{2}[\overline{u} - u] - \underline{c}\right] - \frac{1-\delta\frac{1+q}{2}}{1-\delta q}\left[\overline{c} - \underline{c}\right], \text{ and}$$
$$\overline{\beta}[\overline{u} - \underline{u}] \equiv \overline{p}[\overline{u} - u] - \overline{c} + \frac{\delta}{1-\delta}\left[\frac{1-\overline{p}}{2}[\overline{u} - u] - \underline{c}\right] + \frac{\delta(1-q)}{1-\delta}\frac{p - \underline{p}}{2}[\overline{u} - u].$$

Fix any equilibrium σ^*. First, note that since $\beta \geq 0$, under plurality as under proportional representation, (1) implies that maintaining an existing party is uniquely stage optimal in preference state s_0 for activist j, irrespective of whether activist $-j$ is represented by a party. Hence, by the arguments in the proof of Proposition 1, $\sigma_j^*(s_0, \phi) = 1$ whenever $j \in \phi$. Second, since $\alpha \geq 0$, (1) also implies that $\sigma_j^*(s_j, \phi) = 1$ whenever $j \in \phi$. Third, since no new party faces entry penalty β following entry when $\phi = \emptyset$, (1) also ensures that $\sigma_j^*(s, \emptyset) = 1$ is uniquely optimal when $s \in \{s_0, s_j\}$.

Now consider state $(s_0, \{-j\})$ and equilibrium σ^*. If activist j does not form a party, its payoff is

$$\frac{1+\overline{p}}{2}u + \frac{1-\overline{p}}{2}\underline{u} + \delta\mathbb{E}V_j(s', \{-j\}; \sigma^*),$$

while if activist j forms a party, its payoff is

$$\left(\frac{1-\overline{p}}{2} - \beta\right)\overline{u} + \overline{p}u + \left(\frac{1-\overline{p}}{2} + \beta\right)\underline{u} - \overline{c} + \delta\mathbb{E}V_j(s', \{-j, j\}; \sigma^*).$$

Hence, activist j does not form a party if and only if

$$\overline{c} - \left[\frac{1-\overline{p}}{2}[\overline{u} - u] - \beta[\overline{u} - \underline{u}]\right] \geq \delta\mathbb{E}\Big[V_j(s', \{-j, j\}; \sigma^*) - V_j(s', \{-j\}; \sigma^*)\Big]$$
$$\equiv \delta\mathbb{E}\Delta V_j(s'; \sigma^*) \quad (8)$$

Consider state (s_{-j}, ϕ) such that $j \in \phi$ and such that $\sigma_{-j}^*(s_{-j}, \phi) = 1$. If activist j maintains its party, its payoff is

$$(\underline{p} - \alpha + \beta\mathbb{I}_{-j\notin\phi})\overline{u} + pu + (\overline{p} + \alpha - \beta\mathbb{I}_{-j\notin\phi})\underline{u} - \underline{c} + \delta\mathbb{E}V_j(s', \{-j, j\}; \sigma^*),$$

while if activist j disbands its party, its payoff is

$$(1-\overline{p})u + \overline{p}\underline{u} + \delta \mathbb{E} V_j(s', \{-j\}; \sigma^*).$$

Hence, under profile $\underline{\sigma}^{PL}$, it must be that

$$\underline{c} - \underline{p}[\overline{u} - u] + (\alpha - \beta)[\overline{u} - \underline{u}] \geq \delta \mathbb{E} \Delta V_j(s'; \underline{\sigma}^{PL}), \tag{9}$$

while under profile $\overline{\sigma}^{PL}$, it must be that

$$\underline{c} - \underline{p}[\overline{u} - u] + \alpha[\overline{u} - \underline{u}] \leq \delta \mathbb{E} \Delta V_j(s'; \overline{\sigma}^{PL}). \tag{10}$$

Fix a state (s_j, ϕ) such that $j \notin \phi$. Under $\underline{\sigma}^{PL}$, (1) ensures that the stage payoffs of activist j are strictly positive when it forms a party, so that, by an argument in the proof of Proposition 1, $\underline{\sigma}^{PL}(s_j, \phi) = 1$ is optimal. Under $\overline{\sigma}^{PL}$, activist j forms a party in state (s_j, ϕ) with $j \notin \phi$ if and only if

$$\overline{p}[\overline{u} - u] - \overline{c} + (\alpha - \beta)[\overline{u} - \underline{u}] \geq -\delta \mathbb{E} \Delta V_j(s'; \overline{\sigma}^{PL}). \tag{11}$$

Note that (9), along $\underline{\sigma}^{PL}_{-j}(s_{-j}, \emptyset) = 1$ and the fact that $\overline{c} > \underline{c}$, implies that $\underline{\sigma}^{PL}_j(s_{-j}, \emptyset) = 0$ is optimal. Since the profile $\overline{\sigma}^{PL}$ is specified in all states except $(s_{-j}, \emptyset)$, a simple computation verifies whether either $\overline{\sigma}^{PL}_j(s_{-j}, \emptyset) = 0$ or $\overline{\sigma}^{PL}_j(s_{-j}, \emptyset) = 0$ are optimal. Actions in this state are irrelevant when verifying equilibrium incentives, since under $\overline{\sigma}^{PL}$ it can be reached only following deviations by two activists.

Hence, the relevant incentive constraints under $\underline{\sigma}^{PL}$ are (8) and (9), while the relevant incentive constraints under $\overline{\sigma}^{PL}$ are (8), (10) and (11). These can be further simplified through computation. First, note that

$$\begin{aligned}
\Delta V_j(s_j; \overline{\sigma}^{PL}) &= \overline{c} - \underline{c} + \beta[\overline{u} - \underline{u}], \\
\Delta V_j(s_j; \underline{\sigma}^{PL}) &= \overline{c} - \underline{c}, \\
\Delta V_j(s_{-j}; \underline{\sigma}^{PL}) &= 0,
\end{aligned}$$

so that we have that

$$\Delta V_j(s_{-j}; \overline{\sigma}^{PL}) = \frac{1}{1 - \delta \frac{1-q}{2}} \left[\underline{p}[\overline{u} - \underline{u}] - \alpha[\overline{u} - \underline{u}] - \underline{c} + \delta q \Delta V_j(s_0; \overline{\sigma}^{PL}) + \delta \frac{1-q}{2} \Delta V_j(s_j; \overline{\sigma}^{PL}) \right],$$

and that

$$\Delta V_j(s_0;\overline{\sigma}^{PL}) = \frac{1}{1-\delta q}\left[\frac{1-\overline{p}}{2}[\overline{u}-u]-\underline{c}+\delta\frac{1-q}{2}\Delta V_j(s_{-j};\overline{\sigma}^{PL})+\delta\frac{1-q}{2}\Delta V_j(s_j;\overline{\sigma}^{PL})\right].$$

Further computation yields that

$$\begin{aligned}\delta\mathbb{E}\Delta V_j(s';\overline{\sigma}^{PL}) = \frac{1}{1-\delta\frac{1+q}{2}}&\left[\delta\frac{1-q}{2}\left[\underline{p}[\overline{u}-u]-\alpha[\overline{u}-\underline{u}]-\underline{c}\right]\right.\\ &\left.+\delta q\left[\frac{1-\overline{p}}{2}[\overline{u}-u]-\underline{c}\right]+\delta\frac{1-q}{2}\left[\overline{c}-\underline{c}+\beta[\overline{u}-\underline{u}]\right]\right].\end{aligned}$$

Similarly,

$$\Delta V_j(s_0;\underline{\sigma}^{PL}) = \frac{1}{1-\delta q}\left[\frac{1-\overline{p}}{2}[\overline{u}-u]-\underline{c}+\delta\frac{1-q}{2}\Delta V_j(s_j;\underline{\sigma}^{PL})\right],$$

and further computation yields that

$$\delta\mathbb{E}\Delta V_j(s';\underline{\sigma}^{PL}) = \frac{1}{1-\delta q}\left[\delta q\left[\frac{1-\overline{p}}{2}[\overline{u}-u]-\underline{c}\right]+\delta\frac{1-q}{2}\left[\overline{c}-\underline{c}\right]\right].$$

Evaluated at $\overline{\sigma}^{PL}$, (8) can be rewritten as

$$\beta[\overline{u}-\underline{u}] \geq \frac{1-\delta\frac{1-q}{2}}{1-\delta q}\left[\frac{1-\overline{p}}{2}[\overline{u}-u]-\underline{c}\right]-[\overline{c}-\underline{c}]+\frac{\delta\frac{1-q}{2}}{1-\delta q}\left[\underline{p}[\overline{u}-\underline{u}]-\alpha[\overline{u}-\underline{u}]-\underline{c}\right], \tag{12}$$

while evaluated at $\underline{\sigma}^{PL}$, it can be rewritten as

$$\beta[\overline{u}-\underline{u}] \geq \frac{1}{1-\delta q}\left[\frac{1-\overline{p}}{2}[\overline{u}-u]-\underline{c}\right]-\frac{1-\delta\frac{1+q}{2}}{1-\delta q}\left[\overline{c}-\underline{c}\right]. \tag{13}$$

A straightforward computation verifies that, for any α, the righthand side of (13) is strictly larger than the righthand side of (12), so that (12) holds whenever (13) holds.

Also, (9) can be rewritten as

$$\alpha[\overline{u}-\underline{u}] \geq \underline{p}[\overline{u}-u]-\underline{c}+\beta[\overline{u}-\underline{u}]+\frac{1}{1-\delta q}\left[\delta q\left[\frac{1-\overline{p}}{2}[\overline{u}-u]-\underline{c}\right]+\delta\frac{1-q}{2}\left[\overline{c}-\underline{c}\right]\right], \tag{14}$$

while (10) can be rewritten as

$$\alpha[\overline{u}-\underline{u}] \leq \underline{p}[\overline{u}-u]-\underline{c}+\frac{1}{1-\delta q}\left[\delta q\left[\frac{1-\overline{p}}{2}[\overline{u}-u]-\underline{c}\right]+\delta\frac{1-q}{2}\left[\overline{c}-\underline{c}+\beta[\overline{u}-\underline{u}]\right]\right]. \tag{15}$$

Finally, since the righthand side of (11) is increasing in α, it can be shown by computation to hold for all α if and only if

$$\beta[\overline{u}-\underline{u}] \leq \overline{p}[\overline{u}-u] - \overline{c} + \frac{\delta}{1-\delta}\left[\frac{1-\overline{p}}{2}[\overline{u}-u] - \underline{c}\right] + \frac{\delta(1-q)}{1-\delta}\frac{p-\underline{p}}{2}[\overline{u}-u], \tag{16}$$

That (13) holds follows since $\beta \geq \underline{\beta}$, and that (16) holds follows since $\beta \leq \overline{\beta}$. Hence, conditions (13) and (14) are sufficient for $\underline{\sigma}^{PL}$ to be an equilibrium, while (13), (15) and (16) are sufficient for $\overline{\sigma}^{PL}$ to be an equilibrium. Let $\check{\alpha}$ be the unique value of α such that (14) holds as an equality and define $\underline{\alpha} = \max\{\min\{\underline{p}-\beta, \check{\alpha}\}, 0\}$. Similarly, let $\hat{\alpha}$ be the unique value of α such that (15) holds as an equality and define $\overline{\alpha} = \min\{\max\{0, \hat{\alpha}\}, \underline{p}-\beta\}$. Hence, given any β satisfying (13), $\underline{\sigma}^{PL}$ is an equilibrium if $\alpha > \underline{\alpha}$, while $\overline{\sigma}^{PL}$ is an equilibrium if $\alpha < \overline{\alpha}$. These are sufficient conditions only, since our definition of $\underline{\alpha}$ and $\overline{\alpha}$ embeds the cases when these equilibria fails to exits. Furthermore, (14) and (15) imply that $\underline{\alpha} \geq \overline{\alpha}$, where the inequality is strict whenever $\underline{\alpha}, \overline{\alpha} \in (0, \underline{p}-\beta)$. □

B Appendix: Tables

Table 1: Data

$G_c < 0.05$	$0.05 \leq G_c < 0.10$	$G_c \geq 0.10$
Belgium, Costa Rica, Cyprus, Finland, Germany, Ireland, Israel, Luxembourg, Malta, Netherlands, Norway, Portugal, Romania, Slovakia, Slovenia, South Africa, Sweden, Switzerland, Venezuela, West Germany	Austria, Bermuda, Bosnia-Herzegovina, Bulgaria, Croatia, Estonia, Iceland, Italy, Latvia, Lithuania, Moldova, Niger, Spain	Albania, Australia, Bolivia, Botswana, Canada, Czech Republic, Czechoslovakia, France, Greece, Hungary, Indonesia, Malaysia, Mauritius, Mexico, New Zealand, Poland, Russia, Trinidad and Tobago, Turkey, UK, USA
220 Elections	101 Elections	197 Elections

Notes: All variables are constructed from the Constituency-Level Elections (CLE) Dataset.

Table 2: Static Tests of Duverger's Law

	Total Number of Parties			
Variable	(1)	(2)	(3)	(4)
Average Gallagher Disproportionality Index	-2.58 (4.51)	-2.61 (3.56)	-2.73 (3.81)	-6.54 (3.47)
Decade Dummies Included?	N	Y	Y	Y
Regional Dummies Included	N	N	Y	Y
Flexibly Controlled for Number of Districts?	N	N	N	Y
R^2	0.03	0.05	0.04	0.33
Number of Observations	518	518	518	518

Notes: All variables are specified in logarithms. In particular, each variable x is transformed as $\log(1+x)$. Any party that receives a vote share over 0.05 in any election is counted in the total number of parties. The Average Gallagher Dispropotionality Index for a given country is constructed by averaging the Gallagher Disproportionality Index for each election in the sample for each country. Flexible control for the number of districts is achieved by including sixth order polynomials in the number of districts and in the log-number of districts. Heteroskedasticity robust standard errors clustered by country are presented in parentheses. ** - 99% significance level, * - 95% significance level.

Table 3: Dynamic Tests of Duverger's Law

	Total Entries				Total Exits				Total Net Movements			
Variable	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Average Gallagher Disproportionality Index	-0.74*	-0.88**	-0.79*	-1.05**	-0.54	-0.69*	-0.65*	-0.89**	-0.95	-1.15**	-1.10*	-1.40**
	(0.33)	(0.29)	(0.34)	(0.28)	(0.40)	(0.31)	(0.35)	(0.28)	(0.48)	(0.40)	(0.46)	(0.37)
Decade Dummies Included?	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Regional Dummies Included	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Flexibly Controlled for Number of Districts?	N	N	N	Y	N	N	N	Y	N	N	N	Y
R^2	0.01	0.13	0.16	0.18	0.01	0.10	0.13	0.18	0.01	0.13	0.15	0.19
Num. of Obs.	518	518	518	518	518	518	518	518	518	518	518	518

Notes: All variables are specified in logarithms. In particular, each variable x is transformed as $\log(1+x)$. Entries and exits are computed according to equations (6) and (7) respectively. Total net movements = entries + exits. Flexible control for the number of districts is achieved by including sixth order polynomials in the number of districts and in the log-number of districts. Heteroskedasticity robust standard errors clustered by country are presented in parentheses. ** - 99% significance level, * - 95% significance level.

Table 4: Dynamic Tests of Duverger's Law: Robustness Checks

	Total Entries				Total Exits				Total Net Movements			
Variable	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Average Gallagher Disproportionality Index	–	–	-1.65** (0.55)	-5.11** (1.26)	–	–	-1.39** (0.49)	-5.16** (1.38)	–	–	-3.04** (1.02)	-5.38** (1.10)
Gallagher Disproportionality Index of First Election in Sample	-1.13** (0.28)	–	–	–	-0.95** (0.35)	–	–	–	-1.50** (0.42)	–	–	–
Average District Magnitude	–	0.12** (0.03)	–	–	–	0.11** (0.03)	–	–	–	0.18** (0.04)	–	–
Variables Specified Linearly?	Y	N	N	N	Y	N	N	N	Y	N	N	N
Variables Specified in Logarithms?	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Dep. Var. Constructed With Weights?	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	N
Decade Dummies Included?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Regional Dummies Included	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Flexibly Controlled for Number of Districts?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.19	0.23	0.21	0.50	0.19	0.22	0.21	0.48	0.19	0.24	0.22	0.49
Num. of Obs.	518	544	518	518	518	544	518	518	518	544	518	518

Notes: Total net movements = entries + exits. Average district magnitude = total representatives / number of districts. Each variable x specified in logarithms is transformed as $\log(1+x)$. Dependent variables constructed with weights follow equations (6) and (7), those without weights utilize with unit weights. Flexible control for the number of districts is achieved by including sixth order polynomials in the number of districts and in the log-number of districts. Heteroskedasticity robust standard errors clustered by country are presented in parentheses. ** - 99% significance level, * - 95% significance level.