Advance Access publication November 2, 2011

Political Analysis (2012) 20:235–247
doi:10.1093/pan/mpr041

A Statistical Model for Party-Systems Analysis

Arturas Rozenas
Political Science Department, Duke University, 326 Perkins Building, Durham, NC 27708
e-mail: ar71@duke.edu

Edited by R. Michael Alvarez

Empirical researchers studying party systems often struggle with the question of how to count parties. Indexes of party system fragmentation used to address this problem (e.g., the effective number of parties) have a fundamental shortcoming: since the same index value may represent very different party systems, they are impossible to interpret and may lead to erroneous inference. We offer a novel approach to this problem: instead of focusing on index measures, we develop a model that predicts the entire distribution of party vote-shares and, thus, does not require any index measure. First, a model of party counts predicts the number of parties. Second, a set of multivariate t models predicts party vote-shares. Compared to the standard index-based approach, our approach helps to avoid inferential errors and, in addition, yields a much richer set of insights into the variation of party systems. For illustration, we apply the model on two data sets. Our analyses call into question the conclusions one would arrive at by the index-based approach. Software is provided to implement the proposed model.

1 Introduction

Electoral competition takes a great variety of forms. Polities differ in the number of parties that compete for office, the relative strength of competitors, and the political importance of minor parties. This variety has been an object of voluminous literature in institutional theory and comparative politics (Duverger 1954; Lipset and Rokkan 1967; Riker 1982; Cox 1997; Clark and Golder 2006). What explains variation in the size of party systems? Why is electoral competition sometimes dominated by one party, whereas in other instances, many parties are equally competitive?

A key question in empirical analysis of these questions is "How to count the parties?" In a typical setting, this is not a simple question to answer. Consider the most recent elections in the United Kingdom. The two largest parties, the Conservatives and the Labour, shared two-thirds of the votes. The third largest party, the Liberal Democrats, received one-fourth of the votes. In addition, there were almost 50 other parties whose support varied from a few hundred to almost 1 million votes. Which parties should one count as "relevant"?

A standard practice in dealing with this issue is to use an *index* of party-system fragmentation, which weights parties according to their relative sizes. For example, the most famous of such indexes, "the effective number of parties" (ENP) (Laakso and Taagepera 1979), is the reciprocal of the sum of squared vote-shares. This way, the index discounts smaller parties at the expense of larger ones. Other indexes are designed in a similar fashion (Rae 1967; Molinar 1991).

Although this indexing approach is attractive in its simplicity, it has an irreparable flaw: the same value of an index can represent substantively very different party systems. Accordingly, when index measures of fragmentation are used in a regression analysis, they may lead to erroneous inference: an analysis might suggest that x has an effect on party systems while the true effect is zero (or vice versa). For the very same reasons, the predictions of a model that uses index measures of fragmentation are impossible to interpret in a politically meaningful way.

Author's note: I am grateful Emerson Niou, Iulian Pruteanu-Malinici, Matthew Singer, and two anonymous reviewers for useful comments and suggestions. All errors are mine. Supplementary materials for this article are available on the *Political Analysis* Web site.

© The Author 2011. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

Researchers have been expressing their unease about this property of the ENP (Grofman, Bowler, Blais 2009, p. 6), but no viable solution has been offered to date. Methodologists have been approaching this problem by offering alternative indexes of party system size (Molinar 1991; Dunleavy and Boucek 2003; Golosov 2010). However, this approach is not effective since no one-dimensional measure can adequately represent multidimensional data on party systems; thus, each alternative index measure will face the same set of problems as the ENP.

Hence, in this paper, we move away from summary indexes to propose a new approach to the analysis of party systems. We develop a statistical model that does not require any index measure. The model predicts the *entire* distribution of party vote-shares instead of its one-dimensional summary measure. In this way, the problem that the index measures face (what kind of party system does a given index value represent?) does not even arise. The proposed model prevents inferential fallacies that are easy to make when index measures are used. In addition, such a model has a meaningful interpretation and yields a rich set of insights into variation of party systems.

A party system is a multifaceted entity and can be studied from various perspectives. Recent methodological literature has focused, among other things, on party system volatility (Mustillo 2009), ideological compactness (Alvarez and Nagler 2004), and polarization (Rehm and Reilly 2010). By contrast, the size of a party system has received much less attention in the literature beyond the above indexing approaches. Our treatment of the data on party systems draws attention to their multidimensional nature, thereby contributing to the wider debate in the recent methodological literature on the value of multi-dimensional measures in understanding complex social and political phenomena (Selway 2011).

It is important to specify the relationship between the model presented in this paper and an important class of models of multiparty elections (Katz and King 1999; Honaker, Katz, King 2002; Tomz, Tucker, Wittenberg 2002; Jackson 2002). Although our model for party systems' analysis shares some formal similarities with these models of multiparty elections, they have very different purposes. The multiparty election models were devised to approach research questions like "How does a variable x affect vote-shares of the British Labour, Conservative, or Liberal party?" Unfortunately, these models cannot be used to study questions as "How does a variable x affect the number of competitive parties across the British electoral districts?" or "How does x affect the size of party systems across the nations?" Indeed, since the multiparty election models label parties by their names, they cannot be applied in a cross-national setting because different countries have different sets of parties.

Many theories in comparative politics make predictions about party systems and not particular parties. To take one example, the Duverger's law predicts that under plurality system, there will be two "serious" political parties (Duverger 1954). For such theory, the only relevant information is whether the two largest parties have similar vote-shares and whether the remaining parties, if they exist, have substantially smaller vote-shares. Consider three possible distributions of votes in Table 1. If one is interested in testing the Duverger's prediction, the three distributions of vote-shares ought to be treated as equivalent since they all have only two competitive parties. However, these distributions are clearly different if one is interested in vote-shares of particular parties. Beyond the above-mentioned indexing approaches, the literature has not yet offered a rigorous method for researchers interested in characteristics of *party systems* as opposed to the electoral fortunes of particular parties.[1]

The paper proceeds as follows: First, I briefly discuss the main shortcoming of the index measures of fragmentation, focusing on the most widely used measure, the ENP. Second, I present a new statistical model for analysis of party systems. Third, the model is applied to Japanese electoral data and to one cross-national data set to address the classic question in comparative politics: How does district magnitude and social diversity affect party systems?

The paper is accompanied by an open source software package "ocomposition," which implements the proposed methodology and is available for download at the Comprehensive R Archive Network (CRAN): http://CRAN.R-project.org/package=ocomposition.

[1]Other theories that make predictions about party systems include the $M+1$ theory by Cox (1997) or theories of spatial competition with endogenous number of parties (de Palma, Hong, and Thisse 1990; Myerson and Weber 1993; Osborne and Tourky 2008).

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

Table 1 An example of three distributions of votes that are equivalent if the focus of the research is party systems and not the vote-shares of particular parties

	D1	*D2*	*D3*
Party A	50	50	0
Party B	50	0	50
Party C	0	50	50

2 Revising the ENP

The most popular measure of party system fragmentation is ENP, introduced by Laakso and Taagepera (1979). ENP is the reciprocal of the Herfindahl–Hirschman index used to study market concentration and is defined as $1/\sum_i v_i^2$, where $(v_1, \ldots, v_n)$ are the vote-shares of parties. ENP is a function of an older but less widely used index of fractionalization by Rae (1967), defined as $1 - \sum_i v_i^2$.

The ENP is a widely criticized measure. For instance, Molinar (1991) argued that the ENP can exaggerate party fragmentation: a high ENP can be associated with a party system where one political party dominates others. Other critics have pointed out other properties of party systems that the ENP may either hide or fail to represent (Rae 1967; Bogaards 2004; Dunleavy and Boucek 2003; Golosov 2010). The general problem with the ENP (or in fact any other scalar measure of fragmentation) is that the same ENP value can represent dramatically different party systems.

To see just how significant this problem may be, consider Fig. 1. Here, each distribution represents a party system with three effective parties. The differences between these party systems are indeed striking: The first party system is that of three equally strong parties. The second party system would conventionally be called a two-party system with a few smaller parties. The third party system has one clearly dominant party with a majority of votes. There is really nothing common among these three party systems; yet, they all have the same number of effective parties.

This lack of correspondence between the ENP values and the party systems that they represent has serious implications for empirical research. In particular, a regression analysis that uses the ENP (or other indexes) cannot be reliably employed to either verify or falsify theories of party competition. Suppose a regression analysis estimates that a variable x has no statistically significant effect on the ENP. Does this serve as the evidence that x has no effect party systems? There are two possibilities: either the effect indeed is zero or it could not be identified because the ENP does not discriminate well between very different party systems. One simply cannot know which is the case. Thus, when the ENP is used, many good hypotheses can be falsely rejected and many bad hypotheses can be falsely accepted.

Consider an example from the literature. Cox (1997) has estimated that decreasing the number of seats in a district from 10 to 1 reduces the number of parties "by almost a full (effective) party to 2.09" (p. 214). The result is then interpreted as substantially important because it corresponds to a difference between

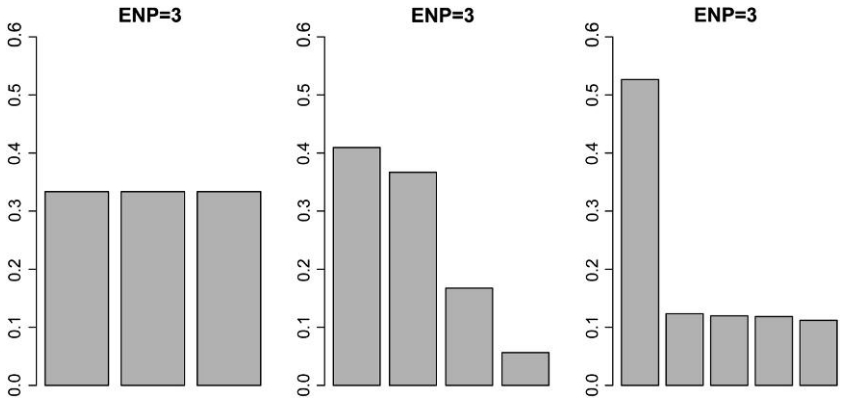


Fig. 1 In all three cases, the ENP is the same, but the party systems that the index is supposed to represent are enormously different.

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

"essentially a two-party legislative system with mostly single-party governments to a two-and-a-half or three-party legislative system with coalition governments as a norm" (p. 217). This interpretation of the analysis is wanting for two reasons.

First, a party system with more than two effective parties can be a system where coalitions are unavoidable *or* a system where coalitions are not at all necessary (compare the first and the third panels in Fig. 1). Second, two effective parties can represent a system, which in no way resembles a two-party system. Suppose the votes are distributed as follows: (0.67, 0.2, 0.08, 0.05). Such a party system would have two effective parties (ENP = 2.01). Yet, it would clearly be better characterized as that with one dominant party. Thus, the ENP measure yields a noninterpretable result. We *cannot* know what a reduction "by a full effective party" means in political terms: it may mean increasing competition among multiple parties or it might mean increasing dominance by a single party. Thus, we cannot use this analysis either to support or to reject the claim that district magnitude increases party system size.

If the ENP has these serious flaws, what is the alternative? Scholars so far have been focusing on devising alternative scalar measures of party system fragmentation (Molinar 1991; Dunleavy and Boucek 2003; Golosov 2010). Unfortunately, since these alternative indexes are also scalar measures, they, by definition, suffer from the same basic flaw as the ENP. It is understandable why so much focus has been paid to the index measures: they provide an easy way to conduct statistical analysis using simple statistical tools as, for instance, linear regression. Simple methods require simple measures; thus, multidimensional data on party systems is reduced to one-dimensional indexes. The above examples should make it clear that such a reduction results in significant loss of information and misleading scientific inference. With advancements in statistical modeling, there is little reason to focus on scalar measures. Instead of trying to find the "holy grail" index of party-system fragmentation, the field would gain more by developing statistical models that represent the inherent multidimensionality of data on party systems.

3 Model

Let $i = 1, \ldots, N$ denote the unit of analysis (e.g., a country or a district). Let $\mathbf{v}_i = (v_{i1}, \ldots, v_{in_i})$ denote the vote-shares of parties, where $\sum_j v_{ij} = 1$. Here, $n_i \in [1, 2, \ldots)$ is the number of parties (or candidates) in a unit i. The vote-shares of parties are arranged in descending order: $v_{i1} \geqslant v_{i2} \geqslant \ldots v_{in_i} > 0$. In this notation, if we have v_{ij}, then the first index i refers to the unit of analysis and the second index j refers to the *rank* of the party. Such ordering of vote-shares is necessary if one is interested in comparing party systems irrespective of the parties' names. Therefore, in our model, we label parties by their rank and not by their name as is done in Katz and King (1999). In sum, $(n_i, \mathbf{v}_i)$ for $i = 1, ..., N$, such that $\sum_j v_{ij} = 1$ and $v_{i1} \geqslant v_{i2} \geqslant \ldots v_{in_i} > 0$ constitutes the data on party systems.

Analysis of data that are in the above form raises two major difficulties. First, these data are *compositional*: vote-shares are constrained to the $(0, 1]$ interval, and they must sum up to one. Methods for compositional data analysis have long been developed (Aitchison 1982). Second, the data on party systems are *ordered*: the goal is to predict the vote-shares of parties arranged from the largest to the smallest party (or vice versa). This second feature is what distinguishes the model proposed in this paper from the models of multiparty election data (Katz and King 1999; Jackson 2002; Honaker, Katz, King 2002; Tomz, Tucker, Wittenberg 2002), which label parties by their names. Models for ordered compositional data, to our knowledge, have not yet been developed.

Though the model in this paper is motivated by the data on party vote-shares, its application does not have to be restricted to this domain. The model can be applied in many other contexts where ordered compositional data appear. For instance, it might be of interest to predict the distribution of seats in parliaments (as an alternative to the "effective number of parliamentary parties"), the allocation of portfolios in coalition governments, or the distribution of ethnic groups (labeled by their rank).

3.1 *Choice of Transformation*

A standard approach in compositional data analysis is to transform the compositional bounded variables $\mathbf{v}$ into noncompositional unbounded variables $\mathbf{y}$. Such transformation allows employing the conventional tools of multivariate analysis. We use the following transformation:

$$\mathbf{y} = (y_1, y_2, \ldots, y_{n-1}) = (l(v_2/v_1), l(v_3/v_2), \ldots, l(v_n/v_{n-1})). \quad (1)$$

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

Here, $l(\cdot)$ is a logistic function, $l(x) = \ln\left(\frac{x}{1-x}\right)$. Note that this transformation is well defined only for $v_j > 0$. The vote-shares $\mathbf{v}$ can be obtained from $\mathbf{y}$ by solving the following system of equations:

$$v_{j+1} = v_j l^{-1}(y_j) \quad \forall\ j = 1, \ldots, n-1 \tag{2}$$

$$\text{s. t.} \sum_j v_j = 1.$$

This transformation differs from the conventional one used in compositional data analysis. Following Aitchison (1982), Katz and King (1999) use a transformation $y_j = \ln(v_j/v_J)$, where v_J is the reference party. Here we do not use the Aitchison's transformation because, when applied to ordered data, it leads to a cumbersome model. For instance, using transformation $y_j = \ln(v_j/v_1)$ yields random variables $(y_1, \ldots, y_{n-1})$ that are unbounded *but* ordered: $v_1 \leqslant v_2 \leqslant \ldots v_n$ implies that $y_1 < y_2 < \ldots < y_{n-1}$. However, to apply the standard tools of multivariate analysis, the transformed data must be both unbounded and unordered, which is impossible to attain using the Aitchison's transformation. In contrast, the transformation in (1) yields random variables $\mathbf{y}$ that are both unbounded and unordered, making further analysis of the transformed data relatively easy.

3.2 *Data-Generating Process*

We treat the transformed data $\mathbf{y}$ as an outcome of a two-stage stochastic process: First, a stochastic count process generates the number of parties taking part in the elections. Second, a stochastic multivariate process determines the distribution of the each party's vote-shares. Formally, let $D = \max(n_i) - 1$ (so $D+1$ is the largest number parties in the data set). Then, the model for the number of parties n_i and their transformed vote-shares $\mathbf{y}_i$ is given by

$$n_i \sim \pi_n \tag{3}$$

$$(y_{i1}, \ldots, y_{iD}) \big| n_i \sim \begin{cases} \mathcal{T}_\nu\left(\boldsymbol{\mu}_{i,[1:j]}, \Sigma_{[1:j,1:j]}\right) & \text{for } j < n_i \\ \mathbf{1}(y_{ij} = \text{NA}) & \text{for } j \geqslant n_i. \end{cases} \tag{4}$$

Here, the number of parties, $n_i \in [1, 2, \ldots)$, is assumed to follow a distribution π_n (the particular choice of a distribution is discussed below). Further, the first $n_i - 1$ elements of $\mathbf{y}_i$ follow a multivariate Student's t distribution with ν degrees of freedom, the vector of means $\boldsymbol{\mu}_{i,[1:j]}$ and the covariance matrix $\Sigma_{[1:j,1:j]}$. The remaining y_is are treated as the "missing data" (here the indicator function $\mathbf{1}(\cdot)$ is used to represent a degenerate distribution). Since we observe the vote-share ratios of only $n_i - 1$ parties, the remaining vote-share ratios are missing data.

In our notation, $\boldsymbol{\mu}_{i,[1:j]}$ refers to the first j elements of the "full" vector of means $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iD})$ and $\Sigma_{[1:j,1:j]}$ refers to the submatrix of the "full" $D \times D$ covariance matrix Σ. For example, if $n_i = 2$, then $y_{i1}|n_i$ is from a univariate t distribution with location μ_{i1} and variance $\Sigma_{[1,1]}$, and $(y_{i2}, \ldots, y_{iD})$ are missing values. If, however, $n_i = D + 1$, then $\mathbf{y}_i|n_i \sim \mathcal{T}_\nu(\boldsymbol{\mu}_i, \Sigma)$, and $\mathbf{y}_i$ contains no missing data.

The model in equations (3–4) is a realistically appealing representation of the electoral process: first, a number of parties decide to run in a race; second, voters distribute their votes across those parties. Note that the *marginal* distribution of y_{ij} is a mixture of Student's t distributions: $\pi(y_{ij}) = \sum_{x=2}^{\infty} \Pr(n_i = x-1)\mathcal{T}_\nu(\mu_{ij}, \Sigma_{(j,j)})$, where $\Pr(n)$ is the probability mass function from the model in (3).

Following Katz and King (1999), we model latent vote-shares using a multivariate t distribution. Political methodologists have debated the advantages of using t distribution over the normal distribution (Honaker, Katz, King 2002; Tomz, Tucker, Wittenberg 2002). When the data contain many outliers, the t model may yield different results, compared to the normal model. Since it represents a more general version, we base our model on the multivariate t distribution. If needed, the t model can be easily turned into a normal model by simply setting $\nu = \infty$.

Although conceptually appealing, the model in equations (3–4) poses computation difficulties due to the missing data: depending on the observed n_i, the dimensionality of the multivariate t distributions will vary across the districts, making it difficult to estimate parameters $\boldsymbol{\mu}$ and Σ. A more convenient representation of the model employs latent variables:

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

$$n_i \sim \pi_n \tag{5}$$

$$(z_{i1}, \ldots, z_{iD}) \sim \mathcal{T}_v(\boldsymbol{\mu}_i', \Sigma) \tag{6}$$

$$y_{ij} = \begin{cases} z_{ij} & \text{if } j < n_i \\ \text{NA} & \text{if } j \geqslant n_i. \end{cases} \tag{7}$$

Here, $(z_{i1}, \ldots, z_{iD})$ are partially observed latent data, representing the transformed vote-shares of the parties. That is, for a given n_i, $(z_{i1}, \ldots, z_{i,n_i-1})$ are observed data and $(z_{in_i}, \ldots, z_{iD})$ are unobserved latent data. Equation (7) relates the model of party counts in equation (5) to the incompletely observed data on party vote-shares modeled by equation (6). The advantage of this specification is that, conditional on $\mathbf{z}_i$, $\boldsymbol{\mu}_i$ and Σ are independent of the partially observed data $\mathbf{y}_i$. This way we can use an efficient Gibbs sampling algorithm to estimate the parameters of the model.

To model party counts n_i, we suggest a *truncated* negative binomial distribution, with mean λ_i and dispersion parameter ω: $n_i \sim \mathcal{NB}(\lambda_i; \tau_i, \omega)\mathbf{1}(l_i \leqslant n_i)$. Here, l_i stands for the lower bound of party counts in a district i. We suggest the use of a truncated distribution because, in many settings, it is unreasonable to use a model that ascribes nonzero probability to certain party counts. For example, a researcher could be interested in modeling the distribution of *candidate* vote-shares as a function of the district magnitude M_i. In such a case, it makes sense to assume that $n_i \geqslant M_i$; otherwise, the model could predict that fewer candidates in a district than available seats. In the application section, we give another example where $l_i > 0$ is a reasonable option. Of course, if needed, one can set $l_i = 0$, yielding the usual negative binomial model.

The predicted party systems are related to the covariates of interest by two sets of regressions. First, the mean party count in a district i is given by $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{q}_i)$ for each $i = 1, \ldots, N$, where $\boldsymbol{\beta}$ is a vector of regression parameters, and $\mathbf{q}_i$ is a vector of covariates for the count model. Similarly, the means of the latent variables $\mathbf{z}_i$ are given by $\mu_{ij} = \boldsymbol{\gamma}_j'\mathbf{x}_i$ for each $j = 1, \ldots, D$ and each $i = 1, \ldots, N$. Here, $\boldsymbol{\gamma}_j$ is a vector of regression parameters, and $\mathbf{x}_i$ is a vector of covariates for the party vote-share model. Note that depending on the goals of analysis, $\mathbf{q}_i$ and $\mathbf{x}_i$ might be the same or different. For a data set with K covariates and a maximum number of parties equal to $D + 1$, the model estimates $K \cdot D$ regression parameters γ and a $D \times D$ covariance matrix Σ. In sum, the model is a composition of a truncated negative binomial regression and a set of multivariate t regressions.

3.3 *Hierarchical Priors and Shrinkage*

The model contains two sets of parameters: count model parameters $(\boldsymbol{\beta}, \omega)$ and vote-share model parameters $(\boldsymbol{\gamma}, \Sigma)$. Whereas the first set of parameters can be easily estimated in the maximum likelihood framework, the second set of parameters is difficult to estimate when the data contain districts or countries with many parties (D is large). The estimation requires integrating over the latent variables $\mathbf{z}_i$, which becomes an increasingly difficult task when D is large. This is a usual problem in compositional data analysis, which can be overcome fairly easily in the Bayesian framework (Butler and Glasbey 2008).

There is one more reason why a Bayesian approach is useful in this setting. In many cases, there will be only a few observations to estimate parameters γ_{jk} when j is large. For example, in the Japanese electoral data that we analyze later in the paper, there are only two districts that have more than seven parties. Consequently, it is difficult to estimate parameters, which control the vote-shares of the eighth and ninth largest parties $(\gamma_{8,1}, \ldots, \gamma_{8,K})$. In fact, it might well be the case that, for large j, there are more parameters $\boldsymbol{\gamma}_j$ to be estimated than available observations. The same problem appears in estimating the covariance matrix Σ.

A Bayesian framework provides a remedy to this problem. We can specify a hierarchical prior structure for parameters $(\boldsymbol{\gamma}, \Sigma)$ and allow information borrowing across the parameters (Gelman et al. 2003, Ch. 5). This way the estimates are shrunken toward their common mean: parameters that cannot be estimated precisely (e.g., γ_{jk} when j is large) are moved toward the estimates of the parameters that are estimated with higher precision (e.g., γ_{jk}'s when j is small). This hierarchical model could be seen as a compromise between a model that assumes fully homogeneous effects (a covariate k has the same effect on the vote-shares of each party) and a model that assumes fully heterogeneous effects.

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

It is a standard approach in the literature on hierarchical models to choose conditionally conjugate priors for regression coefficients and shrinkage parameters (Gelman et al. 2003, Ch. 5). Such priors lead to an efficient computation of the posterior distributions. For our model, conditional conjugacy is attained by assuming for each covariate $k = 1, \ldots, K$,

$$(\gamma_{1k}, \ldots, \gamma_{Dk}) | m_k, t_k \sim \mathcal{N}(m_k, t_k) \tag{8}$$

$$\pi(m_k) \propto 1 \tag{9}$$

$$1/t_k \sim \mathcal{G}(a, b). \tag{10}$$

Here, party-specific regression coefficients $(\gamma_{1k}, \ldots, \gamma_{Dk})$ are shrunken toward their common mean m_k, which is estimated from the data. The hyperparameter t_k controls the degree of shrinkage and is also estimated from the data. We assume that each m_k has a flat (improper) prior and each $1/t_k$ is drawn independently from a gamma distribution with mean a/b and variance a/b^2. We choose $a = b = 1e-3$, resulting in a very vague prior distribution for the parameter $1/t_k$. Note that the prior variance for all these parameters is very large and, therefore, the assumed priors can have only a negligible effect on the estimates.

Furthermore, we assume that Σ is from an inverse Wishart distribution: $\Sigma \sim \mathcal{IW}(w_1, w_2\mathbf{I})$. We let $w_1 = h + D + 1$ and $w_2 = hs$. This implies that a priori, $\mathbb{E}(\Sigma|s) = s\mathbf{I}$. In this parameterization, h controls the degree of shrinkage toward the average residual variance, s. It is assumed that a priori (h, s) are from a standard bivariate log-normal distribution. Notice that these are the hyperparameters of the model and so the prior distributions can have only a minor effect on the parameter estimates. Following Liu (1996), the degrees of freedom parameter ν is given a noninformative distribution $\pi(\nu) \propto \nu^{-2}\mathbf{1}(\nu \geqslant 1)$.

Lastly, for the negative binomial regression, we assume noninformative improper priors for the regression parameters, $\pi(\boldsymbol{\beta}) \propto 1$, whereas the dispersion parameter w is given a log-normal distribution with mean equal to 1 and variance equal to 100 (a larger variance does not change the results but slows down the convergence of the sampling algorithm). Marginal posterior distributions of the parameters are computed using Markov chain Monte Carlo methods. The sampling algorithm is described in Section A1 of the supplementary material.

3.4 *Interpretation and Prediction*

How do we judge if a variable x_k has a substantial effect on the party system? Depending on the goals of a study, there are several ways to analyze the output of the model. First, we suggest to look at the count model parameter β_k, which measures the effect of x_k on the average party count, and the vote-share model hyperparameter m_k, which represents the average effect of the covariate x_k on the party vote-shares. If m_k is large and negative, it indicates that as x_k increases, the party system approaches that of highly uneven distribution of votes. Conversely, if m_k is large and positive, then increasing x_k leads to a party system with an even distribution of vote-shares.

In principle, if the estimates of both β_k and m_k are significantly far away from zero, there is strong evidence that x_k has an effect on party systems. The direction of this effect can be evaluated looking at the signs of the parameters. For instance, scholars are often interested in the size of party systems. One could define a party system as "large" if it has many parties *and* if the votes are distributed evenly—there are many competitive parties. In this case, one would say that x increases the size of a party system if β_k and m_k are positive. When β_k is positive, whereas m_k is not, we can say that x_k increases the number of parties but not the size of a party system as the smaller parties are not competitive.

Second, if a researcher wants more precise knowledge as to how x_k affects vote-shares of individual parties, one can study regression coefficients γ_{jk}. If, say γ_{1k} is negative, then increasing x_k, increases the vote-share gap between the first and the second largest parties. Alternatively, one can study how the predicted vote-shares of each party change as a function of x_k.

Third, the most informative way to interpret the results is by setting the explanatory variables to the values of interest and then simulating a predicted party system or simulating several predicted party systems and comparing them. The simulation procedure is described in Section A2 of the supplementary material.

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

When analyzing the predicted party systems, it is important to know whether the difference between two predicted party systems is significant in statistical terms. We provide a simple method to measure the uncertainty about predicted party systems. Let Θ refer to a set of parameters of the data generating process and let $\hat{\mathbf{v}}(\mathbf{x}'|\Theta)$ and $\hat{\mathbf{v}}(\mathbf{x}''|\Theta)$ be two predicted party systems given covariate values $\mathbf{x}'$ and $\mathbf{x}''$, respectively. Let d be the distance between these two party systems defined as

$$d(\mathbf{x}', \mathbf{x}'') = \frac{1}{2}\sum_{i=1}^{D}\left|\hat{v}_i(\mathbf{x}'|\Theta) - \hat{v}_i(\mathbf{x}''|\Theta)\right|. \tag{11}$$

The metric d is equal to zero if the two party systems are exactly the same and it approaches 1 as two party systems become increasingly different. Since d depends on a set of unknown parameters Θ, it is a random variable. We can use the following expression to measure uncertainty about predicted party systems:

$$\pi_\epsilon = \Pr_{\Theta}[d(\mathbf{x}', \mathbf{x}'') > \epsilon]. \tag{12}$$

π_ϵ is the posterior probability (calculated with respect to the joint posterior distribution of all unknown parameters Θ) that the distance between two predicted party systems is at least ϵ, where $\epsilon \in (0, 1)$ is a chosen threshold. For most practical purposes, the distance of less than 0.1 corresponds to essentially equivalent party systems. Thus, we suggest using $\epsilon = 0.1$ in empirical applications.[2] If the posterior probability π_ϵ is small, then the difference between two predicted party systems can be said to be statistically intractable.

4 Applications

One of the most widely studied questions in comparative politics and institutional theory is "How do electoral institutions and social diversity affect party systems?" (Duverger 1954; Ordeshook and Shvetsova 1994; Cox 1997; Clark and Golder 2006). In this section, we apply our model to two data sets. First, we analyze district-level party systems in Japan. Second, we analyze how the average district magnitude affects national party systems across 40 nations. The data and the code used in the analyses are available at http://hdl.handle.net/1902.1/16637.

4.1 *District-level Party Systems in Japan*

Here we analyze the data on district-level party systems in Japan from 1960 to 1993 (Reed n.d.), which were used by Cox (1997) to study the effects of district magnitude and social diversity on party systems. In his study, Cox regressed the ENP on district magnitude, urbanization (a proxy for social diversity), and their interaction (plus the lagged dependent variable and year dummies). A simple regression of the ENP on these covariates leads to the conclusion that both district magnitude and urbanization have statistically significant positive effects on the size of party systems and that these effects are interactive.

Using the same set of covariates,[3] we arrive at somewhat different conclusions. Table 2 reports the summary estimates of the negative binomial regression parameters $\boldsymbol{\beta}$ and ω as well as vote-share model hyperparameters $\mathbf{m}$ and ν. District magnitude and urbanization increase the raw number of parties. However, the effect of these covariates is only additive since the credible interval of the interaction term includes zero.

Furthermore, in the vote-share model, only urbanization has a statistically tractable effect: the coefficient m_u is positive (0.22) and its 95% highest posterior density (HPD) interval borders near zero. Thus, compared to rural districts, urban districts have more parties *and* the votes across all parties are distributed more evenly. In comparison, we do not observe strong evidence that district magnitude increases the size

[2]The choice of ϵ is akin to the choice of a rejection region in the classical hypothesis testing: if ϵ is close to zero, then the hypothesis of no difference will be always accepted ($\pi_\epsilon = 1$); if ϵ is close to one, it will be always rejected ($\pi_\epsilon = 0$).

[3]The only difference is that, in the vote-share model, we also add n_i as the covariate to allow the relative vote-shares to vary with the number of parties. For example, it could be the case that in a district with few parties, votes across these parties are distributed equally only because there are few parties. This addition does not make difference for these particular data but it could make difference when another data set is analyzed; hence, we suggest to use it as a general rule.

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

Table 2 Posterior averages and 95% highest posterior density intervals. The interacted variables are centered around the sample mean. Lagged counts and the year dummies were included in the count model

	Count model		*Vote-share model*	
	$\hat{\beta}$	*95% HPD*	$\hat{m}$	*95% HPD*
Intercept	0.52	[0.34, 0.69]	−0.72	[−1.75, 0.33]
Magnitude, M	0.08	[0.05, 0.11]	−0.14	[−0.34, 0.09]
Urbanization, U	0.08	[0.05, 0.10]	0.22	[−0.01, 0.42]
Magnitude x Urbanization	−0.02	[−0.04, 0.01]	0.04	[−0.21, 0.35]
$ln(\omega)$	14.38	[7.00, 24.92]		
ν			15.59	[13.11, 18.98]
N (districts)		1373 (141)		
Correctly predicted[a]		79%		
Mean absolute error		0.03		

[a]Percentage of correctly predicted party vote-shares at the 5% tolerance level.

of the party system: the coefficient m_M is negative (−0.14) and not statistically significant. Thus, increasing district magnitude encourages entry of new parties but the overall distribution of votes across the parties does not become more even.

We can also look at how district magnitude and urbanization affect each party's vote-share individually. Figure 2 shows the change in predicted vote-shares of the parties as we change each variable by two units (about 2 SDs) holding the other variables at their median values. Increasing urbanization reduces the vote-share of the leading party by about 25%. In comparison, increasing district magnitude does not have a statistically significant effect on the vote-share of the largest party as the 95% HPD interval crosses the zero line. In fact, district magnitude has a statistically significant effect only on the vote-share of the second largest party. This results has an important substantive implication: increasing district magnitude increases the number of parties, which takes away the votes from the second largest party leading to a less competitive party system.

Lastly, Fig. 3 shows *average* predicted party systems given four constellations of the independent variables (again, for both variables we chose 1st and 9th decile in the data). First, consider the upper panel of Fig. 3. In rural constituencies, as district magnitude increases, we observe more parties entering competition. Remarkably, however, the smaller parties *do not* reduce the fortunes of the largest party. Instead, as the number of parties grows, the vote-shares of the smaller parties fall; thus, the largest party *increases* its relative dominance in large-magnitude rural constituencies.

Now consider the lower panel of Fig. 3. Here we see how district magnitude affects party systems in urban districts. Notice, first, the difference between the urban and rural constituencies: in urban areas, party systems are much more competitive as the dominant party does not exist. This is especially visible

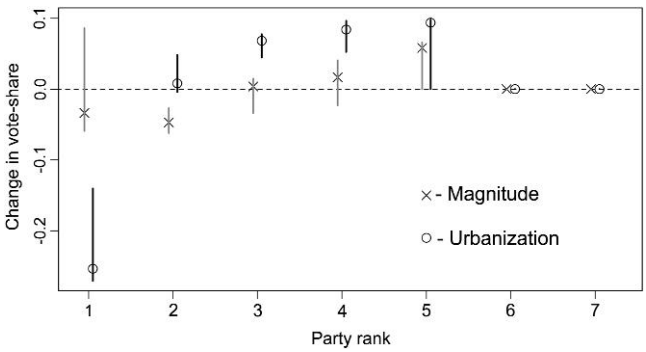


Fig. 2 Median differences in predicted vote-shares with 95% HPD intervals as district magnitude and urbanization change by two units (approximately 2 SDs) while holding other variables at their median values and setting year to 1993.

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

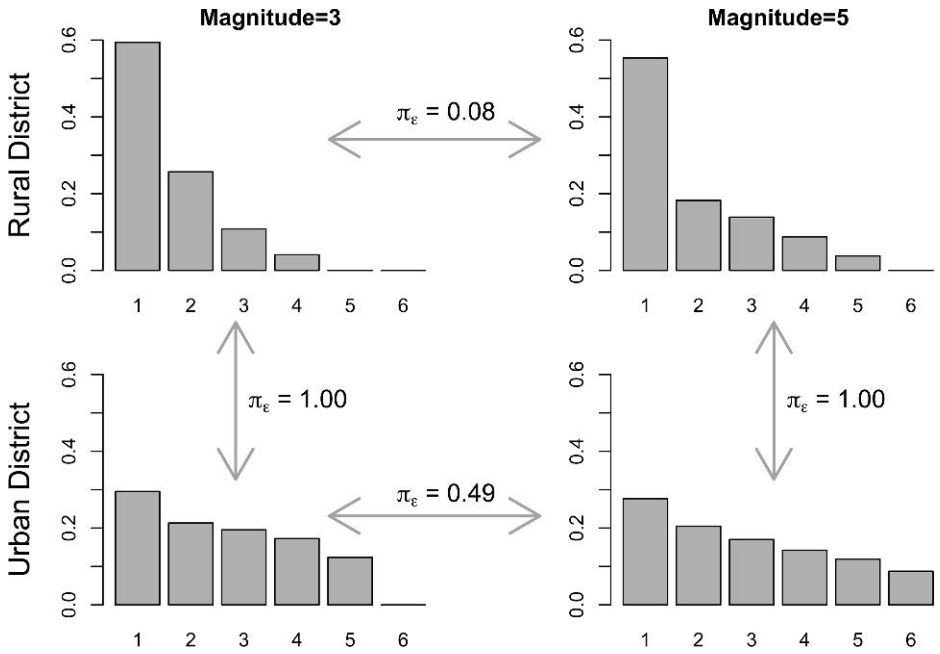


Fig. 3 Average predicted party systems in Japanese electoral districts as a function of district magnitude and urbanization. π_ϵ is the posterior probability that the distance between the two predicted distributions is at least ϵ (here, $\epsilon = 0.1$).

in the lower magnitude districts where vote-shares of multiple parties are very close. However, as the district size increases, two things start to change: there are more parties participating in elections and the distribution of votes becomes less even. Remarkably, there are more parties taking part in elections but the party system becomes *less* competitive.

Figure 3 also shows the posterior probabilities of difference between each predicted party system, π_ϵ. These probabilities also suggest that the effects of district magnitude on party systems are very minor: as we go from a three-seat to a five-seat district, the posterior probability of difference between the predicted party systems (π_ϵ) is very small both in rural (0.08) and in urban districts (0.49). By comparison, the predicted party systems are very different when we go from rural to urban districts as witnessed by very large π_ϵs. As we suggested, the overall effect of district magnitude is minor because it significantly affects only the second largest party.

These results call into question the notion that larger districts should have more competitive parties (Cox 1997). Indeed, larger districts have more parties but not more competitive parties (where by 'competitive' we mean parties with relatively large vote-shares). One possible interpretation of this result is that electoral coordination fails in larger districts: more parties enter competition, their votes get split, and, in turn, this benefits the largest party. On the other hand, in small districts, smaller parties coordinate their behavior more effectively: they are more cautious about entering the competition and thereby splitting their vote, which, interestingly, benefits the second largest party. Thus, electoral coordination seems to be more successful in small districts than in large ones—a result in line with the theoretical argument by Cox (1997).[4]

We can draw several lessons from this analysis. First, larger district magnitude does not lead to a larger party system, contrary to what the conventional analysis using the ENP measure would suggest. Instead, our analysis suggests that larger districts have fewer truly competitive parties. Second, our model generated a set of novel findings about Japanese politics related to the effects of urbanization, the dominance of the largest parties, and the effects of electoral institutions on the fortunes of the second largest party. None of these results could have been generated using the conventional approach to party system analysis.

[4]I am indebted to Matthew Singer for drawing my attention to this point.

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

Table 3 Summary of the posterior estimates for the negative binomial regression parameters and the vote-share model hyper-parameters m

	Count model		*Vote-share model*	
	$\hat{\beta}$	*95% HPD*	$\hat{m}$	*95% HPD*
I	1.79	[1.64, 1.90]	0.09	[−0.06, 0.27]
ln(Magnitude)	0.03	[−0.02, 0.09]	0.09	[−0.01, 0.19]
ln(w)	11.79	[4.31, 22.41]		
ν			2.15	[1.61, 2.79]
Elections (countries)		88 (40)		
Correctly predicted[a]		78%		
Mean absolute error		0.03		

[a]Percentage of correctly predicted party vote-shares at the 5% tolerance level.

4.2 *Party Systems Across Nations*

In this section, we use the Comparative Study of Electoral Systems (www.cses.org) data to study the effect of average district magnitude on national party systems. Our main purpose is to compare the conclusions one would draw from the conventional index-based model and our proposed model. Thus, we include the log of average district magnitude as the only covariate in our analysis. In a simple regression model that uses the ENP as the dependent variable, the estimated effect of ln(Magnitude) is 0.30 with the standard error of 0.07. This implies that going from a country with 1 seat per district to a country with 25 seats per district increases the party system size by a full effective party. This would seem to be strong evidence that electoral institutions have a substantial effect on party system size. However, as we noted before, such a conclusion would be unwarranted given the properties of the ENP measure.

Our model suggests that the effects of the average district magnitude are statistically significant but substantively they are vanishingly small. Table 3 shows that increasing district magnitude increases the number of parties and makes the distribution of party vote-shares more even—the lower bound of the 95% HPD intervals for both β and m border near zero. However, in substantive terms, the effects are very small

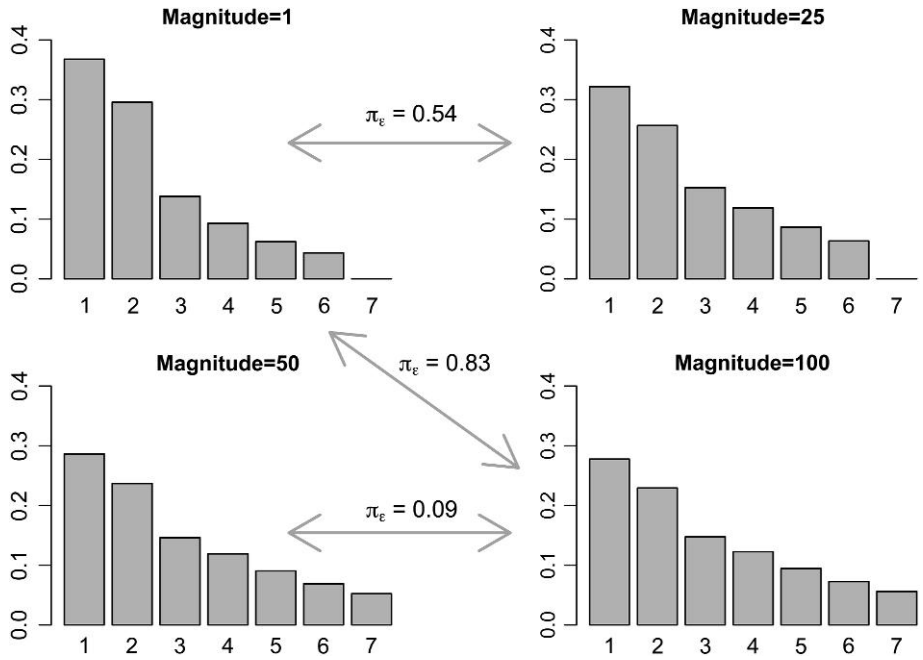


Fig. 4 Average predicted national party systems as a function of average district magnitude. π_ϵ represents the posterior probability that the distance between the two predicted distributions is at least ϵ (here, $\epsilon = 0.1$).

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

as can be seen in Fig. 4. For instance, compare a predicted party with one seat per district to its counterpart with 25 seats: statistically, they are hardly different because the posterior probability π_ϵ is only 0.54. This means that the substantial effect of this change in district magnitude is too small to outweigh the statistical uncertainty of the prediction. In other words, there is a 46% chance that the distance between these two party systems is insubstantial (below 0.1 in this case). Only when we compare the predicted party systems with 1 seat versus 100 seats do we start seeing the effect of district magnitude that is substantial enough to outweigh the statistical uncertainty: here the probability π_ϵ increases to a reasonably large size of 0.84, leading one to believe that given the uncertainty in the estimated parameters, these two predicted party systems are very likely to be different.

This result suggests that the average district magnitude has a very minor effect on the size of a national party system. This effect can only be traceable if we compare extremely small and extremely large districts. When we compare polities with moderately different district magnitudes, the effect is practically untraceable. An analysis that uses the ENP measure would greatly overstate the effect of the average district magnitude on the size of national party systems.

5 Discussion

Political scientists are increasingly recognizing that complex social and political phenomena cannot be studied using simplistic one-dimensional measures Selway (2011). This issue is especially salient in the studies of party systems. Though scholars have long been acknowledging the multidimensional nature of party systems (Sartori 1976; Gross and Sigelman 1984), no systematic method has been developed to analyze party systems statistically. This paper proposed a model that uses the data on party-systems in its original multidimensional form instead of reducing them to an index measure and, in the process of this reduction, losing important information. We have shown that this approach generates interpretable results, helps to avoid errors of inference, and yields a great deal of insight into the variation of party systems.

To illustrate the richness of insights one can get out of the model, we applied it to two different data sets. First, using district-level electoral data from Japan, we have shown that district magnitude does not increase the number of competitive parties in Japanese electoral districts and that the effect of district magnitude is in general much more moderate than the effect of urbanization. We have also identified a very clear pattern of how electoral coordination fails in larger districts: more parties enter competition but their votes get split up, which then benefits larger parties. Second, using cross-national data, we have found that average district magnitude in a country increases its party system size. However, the substantive effect of the average district magnitude is much smaller than the conventional approach of using the index measures would suggest. These analyses were only of illustrative nature and future research could exploit the proposed model in a more extensive study of these issues.

This paper presented a model where ordered compositional data represent dependent variables in the analysis. However, it is often the case that ordered compositional data represent exogenous variables. For example, a big question in comparative politics concerns the effects of ethnic fragmentation on various outcomes (civil conflict, economic growth, or political competition) or the effect of party system size of multiple outcomes of interest. Here, researchers are also using index measures of fragmentation and therefore their results are subject to the interpretational shortcomings and inferential errors that we outlined in this paper. Development of statistical methods for this type of problem would be a significant improvement over the existing approaches.

References

Aitchison, Jim. 1982. The statistical analysis of compositional data. *Journal of the Royal Statistical Society, Series B (Methodological)* 44:139–77.

Alvarez, Michael R., and Jonathan Nagler. 2004. Party system compactness: Measurement and consequences. *Political Analysis* 12:46–62.

Bogaards, Matthijs. 2004. Counting parties and identifying dominant party systems in Africa. *European Journal of Political Research* 43:173–97.

Butler, Adam, and Chris Glasbey. 2008. A latent Gaussian model for compositional data with zeros. *Journal of the Royal Statistical Society* 57:505–20.

Clark, William Roberts, and Matt Golder. 2006. Rehabilitating Duverger's theory. *Comparative Political Studies* 39:679–708.

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press

Cox, Gary W. 1997. *Making votes count*. Cambridge: Cambridge University Press.
de Palma, A., Gap-Seon Hong, and J. F. Thisse. 1990. Equilibria in multi-party competition under uncertainty. *Social Choice and Welfare* 7:247–59.
Dunleavy, Patrick and Françoise Boucek. 2003. Constructing the mumber of parties. *Party Politics* 9:291–315.
Duverger, Maurice. 1954. *Political parties*. New York: Wiley.
Gelman, Andrew, John B. Carlin, Hal B. Stern, and Donald B. Rubin. 2003. *Bayesian data analysis*. 2nd ed. London: Chapman and Hall.
Golosov, Grigorii V. 2010. The effective number of parties. *Party Politics* 16:171–92.
Grofman, Bernard, Shaun Bowler, and Andre Blais. 2009. Introduction: Evidence for Duverger's law from four countries. In *Duverger's law of plurality voting*, eds. Bernard Grofman, Andre Blais, and Shaun Bowler, 1–11. New York: Springer.
Gross, Donald A., and Lee Sigelman. 1984. Comparing party systems: A multidimensional approach. *Comparative Politics* 16:463–79.
Honaker, James, Jonathan Katz, and Gary King. 2002. A fast, easy, and efficient estimator for multiparty electoral data. *Political Analysis* 10:84–100.
Jackson, John E. 2002. A seemingly unrelated regression model for analyzing multiparty elections. *Political Analysis* 10:49–65.
Katz, Jonathan N., and Gary King. 1999. A statistical model for multiparty electoral data. *American Political Science Review* 93(1):15–32.
Laakso, Markku, and Rein Taagepera. 1979. Effective number of parties: A measure with application to West Europe. *Comparative Political Studies* 12:3–27.
Lipset, Seymour Martin, and Stein Rokkan. 1967. Cleavage stuctures, party systems and voter alignments: An introduction. In *Party systems and voter alignments: Cross-national perspectives*, eds. S. M. Lipset and Stein Rokkan. New York: Free Press.
Liu, Chuanhai. 1996. Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association* 91:1219–27.
Molinar, Juan. 1991. Counting the number of parties: An alternative index. *American Political Science Review* 85:1383–91.
Mustillo, Thomas J. 2009. Modeling new party performance: A conceptual and methodological approach for volatile party systems. *Political Analysis* 17:311–32.
Myerson, Roger B., and Robert J. Weber. 1993. A theory of voting equilibria. *American Political Science Review* 87:102–14.
Ordeshook, Peter C., and Olga V. Shvetsova. 1994. Ethnic heterogeneity, district magnitude, and the number of parties. *American Journal of Political Science* 38:100–23.
Osborne, Martin J., and Rabee Tourky. 2008. Party formation in single-issue politics. *Journal of the European Economic Association* 6:974–1005.
Rae, Douglas W. 1967. *The political consequences of electoral laws*. New Haven, CT: Yale University.
Reed, Steven R. n.d. *Japan MMD data set*. www.fps.chuo-u.ac.jp/∼sreed/DataPage.html.
Rehm, Phillip, and Timothy Reilly. 2010. United we stand: Constituency homogeneity and comparative party polarization. *Electoral Studies* 29:40–53.
Riker, William. 1982. The two-party system and Duverger's law: An essay on the history of political science. *American Political Science Review* 76:733–66.
Sartori, Giovani. 1976. *Parties and party systems. A framework for analysis*. Cambridge: Cambridge University Press.
Selway, Joel S. 2011. The measurement of cross-cutting cleavages and other multidimensional cleavage structures. *Political Analysis* 19:48–65.
Tomz, Michael, Joshua A. Tucker, and Jason Wittenberg. 2002. An easy and accurate regression model for multiparty electoral data. *Political Analysis* 10:66–83.

https://doi.org/10.1093/pan/mpr041 Published online by Cambridge University Press