Political Analysis Advance Access published October 30, 2006

doi:10.1093/pan/mpl006

Accounting for Voter Heterogeneity within and across Districts with a Factor-Analytic Voter-Choice Model

Wagner A. Kamakura
Fuqua Graduate School of Business, Duke University,
One Towerview Road, Durham, NC 27708
e-mail: kamakura@duke.edu (corresponding author)

José Afonso Mazzon
Faculdade de Administração e Economia, University of São Paulo,
Ave. Prof. Luciano Gualberto, 908, CEP 0558-900 São Paulo, Brazil
e-mail: jamazzon@usp.br

In this study, we propose a model of individual voter behavior that can be applied to aggregate data at the district (or precinct) levels while accounting for differences in political preferences across districts and across voters within each district. Our model produces a mapping of the competing candidates and electoral districts on a latent "issues" space that describes how political preferences in each district deviate from the average voter and how each candidate caters to average voter preferences within each district. We formulate our model as a random-coefficients nested logit model in which the voter first evaluates the candidates to decide whether or not to cast his or her vote, and then chooses the candidate who provides him or her with the highest value. Because we allow the random coefficient to vary not only across districts but also across unobservable voters within each district, the model avoids the *Independence of Irrelevant Alternatives Assumption* both across districts and within each district, thereby accounting for the cannibalization of votes among similar candidates within and across voting districts. We illustrate our proposed model by calibrating it to the actual voting data from the first stage of a two-stage state governor election in the Brazilian state of Santa Catarina, and then using the estimates to predict the final outcome of the second stage.

1 Introduction

Since the seminal work by Katz and King (1999), there has been a growing interest in modeling and predicting actual voter behavior in multiparty elections, probably motivated by the fact that multiparty elections are the norm rather than the exception throughout the world. Before this new research stream, most of the literature on voting behavior was based on survey data, rather than actual voting behavior or, when based on real votes, utilized national-level data for comparative political analysis. Moreover, the vast majority of prior studies dichotomized the electoral system into an artificial two-party system (Katz and King 1999). With a few exceptions (Glasgow 2001), most of the research extending the work by Katz and King (1999) focuses on the statistical modeling of aggregate voting behavior (Honaker, Katz, and King 2002; Jackson 2002; Tomz, Tucker, and Wittenberg

2002; Mikhailov, Niemi, and Weimer 2002). This might be explained by the fact that actual election results are only available at a certain level of aggregation, to ensure ballot inviolability. Models of voter choice are typically estimated on sample or panel data on self-reported voting or intentions, rather than actual voting behavior.

The approach we propose in this study attempts to bridge the gap between statistical vote-share models (such as those cited above) applied to district-level data, and models of individual voting behavior such as Glasgow's (2001) "mixed" logit model. Starting with the classic spatial theory of voter-choice behavior, we develop a factor-analytic nested logit model for the individual voter's decision to cast a ballot and choose a candidate. We then assume that each district contains a diverse population of voters with a distribution of political preferences to be estimated. By integrating over this distribution of political preferences within each district, we produce a model of individual voting behavior that is consistent with the aggregate data available for each electoral district, and therefore can be estimated at this aggregation level while preserving the theoretical underpinnings of a voter-choice model.

As we will demonstrate later, the model we propose combines several useful features to the political scientist interested in modeling individual voting behavior using actual election data. First, the proposed model considers that abstentions might compete for voters' choices in a different way that candidates compete with each other, so that the elimination of a candidate has different consequences for abstentions than to the remaining candidates. Second, instead of using prespecified and self-reported measures of the candidates' position on issues, typically collected from a small sample and assumed to be relevant to all voters, our model infers a perceptual mapping of the candidates and districts directly from the observed voting data. With this feature, the political analyst has a positioning map of the candidates and districts based only on actual voting behavior, showing how candidates compete with each other for the districts' preferences and how the districts differ in their preferences for the candidates. Most importantly, our model assumes that each district is composed of a diverse population with heterogeneous political preferences, and that districts differ from each other in their average political preferences. This feature is critical for the understanding of voting behavior because it explains how the candidates compete in catering to the needs of a diverse population of voters within each district and across districts. Finally, although our model is developed from individual voter primitives, it is consistent with the aggregate behavior observed at the district level. Therefore, the model can be estimated with district-level data even though it is based on assumptions regarding individual voting behavior.

In the section that follows, we develop our model starting with individual voting behavior and aggregating it to the district level. Next, we show how our model can be estimated using aggregate data and discuss its implications to the political analyst. We then apply our proposed model to data from a two-stage election, where we use the actual district-level results from the first round for parameter estimation and then apply the estimated model to predict the outcome of the second round for each electoral district. We use this illustration to compare the performance of our proposed model with other voter-choice models and to discuss its implications for the political analyst.

2 A Factor-Analytic Nested Logit Voter-Choice Model

To develop our voter-choice model, we start with the classic spatial theory of voting behavior (Black 1958; Downs 1957), assuming that individual voters make their choices through a comparison of their own preferences on issues and policies and their perceptions of the candidates' positions on these issues or policies. We assume that voters do not use their choices in the first stage of the two-stage election strategically, and therefore will cast their ballots on the candidate they perceive as best for them in that election. While developing our model, we first followed the classical "Downsian" ideal-point model where politicians and voters are positioned in the "issues" or "policies" space, and voters select the candidate closest to their own position (Enelow and Hinich 1984). However, our preliminary results showed candidates clustered around the center of the latent issues space and voting districts positioned away from the center, strongly suggesting a vector formulation as more adequate and parsimonious. Therefore, we chose the more recent directional theory (MacDonald, Rabinowitz, and Listhaug 2001), where the voter rewards the candidate who takes a strong position on issues in the direction of the voter's preference (Cho and Endersby 2003).

However, instead of using self-reported perceptual measures of the candidates on predefined issues or policies as predictors of voter choice, we infer a latent issues space, defined by the position of each candidate and the directional preferences for each district, estimated directly from the observed voting behavior. We adopt the latent-variables approach originally proposed by Elrod (1988), letting the actual voting behavior determine the dimensionality and nature of the latent space differentiating politicians relative to voter preferences. This latent-factor formulation is similar to the one used in recent data reduction models (Wedel and Kamakura 2001; Quinn 2004), except that instead of using manifest variables such as self-reported measures or individual characteristics, we use actual voting behavior as indicators for the latent factors.

Following the probabilistic voting literature (Coughlin and Nitzan 1981; Coughlin 1992; Zeng 2000), we assume that voting behavior is affected by many unobservable factors beyond the candidates' position along the voters' preferred directions and observed voter characteristics, leading to a stochastic model (Zeng 2000; Glasgow 2001; Dow and Endersby 2004). Therefore, the value of candidate j to voter n in district i depends on observable and unobservable characteristics of the district and the candidate ($V_{jn(i)}$), as well as a random component ($\varepsilon_{jn(i)}$) that captures the effects of all other factors not explicitly considered in the model:

$$U_{jn(i)} = V_{jn(i)} + \varepsilon_{jn(i)}. \tag{1}$$

The total value of candidate j to a voter n residing in district i is then given by

$$U_{jn(i)} = \alpha_j + \beta_j X_{n(i)} + \lambda_j Z_{n(i)} + \varepsilon_{jn(i)}, \tag{2}$$

where

- α_j = candidate-specific intercept,
- $X_{n(i)}$ = vector of demographic characteristics for voter n's electoral district i,
- β_j = vector of demographic coefficients representing candidate j's appeal to a particular demographic constituency, relative to other candidates,
- $Z_{n(i)}$ = p-dimensional vector of latent scores,
- λ_j = p-dimensional vector of factor weights (or loadings) for candidate j, and
- $\varepsilon_{jn(i)}$= extreme-valued random error.

The first component in the right-end side of equation (2) captures the effect of observed voter characteristics. However, because the characteristics of individual voters are not observable, the estimated coefficients β_j for each candidate refer only to the aggregate characteristics of each district and must therefore be interpreted accordingly. The second component ($\lambda_j Z_{n(i)}$) accounts for unobservable differences in voter preferences within and across voting districts, relative to the population average.

Unobservable differences in voter preferences *within* each district are accounted for by assuming that the p-dimensional[1] latent scores ($Z_{n(i)}$) are independently distributed within each district i with a normal distribution $\phi(\mu_i, \sigma_i)$. We account for unobservable differences in preferences *across* districts by assuming that the mean preferences for each district (μ_i) have standardized normal distributions. This assumption results in standardized scores for each district in the issues space formed by the factor loadings (λ_j), defining the average utility of candidate j to district i. The assumption of a standardized normal distribution for the mean preferences does not imply any loss of generality, as differences in variances in mean preferences across districts will be captured by the factor loadings.

Even though we assume the random errors $\varepsilon_{jn(i)}$ to be independent, the random utilities $U_{jn(i)}$ are correlated across voters and candidates; this correlation is captured by the factor structure ($\lambda_j Z_{n(i)}$), as shown by Wedel and Kamakura (2001). Because of this correlation of the random utilities for the competing candidates, the proposed model will avoid the proportional-draw assumption implicit in the popular multinomial logit model because random utilities for candidates that are positioned closer to each other in the latent space will have higher correlations, thus accounting for the cannibalization of votes among similar candidates. Accounting for cannibalization among similar candidates is particularly important when modeling the choice behavior in two-stage elections[2] because typically a large number of candidates in the first stage is reduced to only a few candidates in the second stage, and it is possible that the eliminated candidates "cannibalized" votes from one of the remaining candidates in the first stage, as we will show later in our empirical application. Most importantly, because we also allow for unobserved voter heterogeneity within each district, the model will also account for the potential cannibalization of votes among similar candidates, which might happen within each district, given that the district may aggregate a diverse population of voters.

If a voter $n(i)$ casts a ballot, the probability that he or she will vote for candidate j is given by (McFadden 1974):

$$P_{jn(i)|V} = \frac{e^{\alpha_j + \beta_j X_i + \lambda_j Z_{n(i)}}}{\sum_{j'} e^{\alpha_{j'} + \beta_{j'} X_i + \lambda_{j'} Z_{n(i)}}}. \tag{3}$$

The model described so far is similar to the random-coefficients or mixed multinomial logit model for multiparty elections proposed by Glasgow (2001), which accounts for differences in preferences among individual voters. Like Glasgow's model, we allow voters to vary in their preferences, assuming a normal distribution of response coefficients ($Z_{n(i)}$) across voters n and districts i. However, our proposed model has two important differences from Glasgow's mixed multinomial logit model. First, instead of using self-reported measures of voter perceptions of the candidates, we estimate the position of the candidates in a latent space (λ) directly from the voting data, providing the political

[1]The dimensionality p of the latent space is determined empirically, as shown later.

[2]Two-stage elections are prevalent across the world. In this system, a large number of candidates compete in the first round. If the winner does not obtain a minimum proportion of the votes, a second round of elections is run between the winner and runner-up.

scientist with a positioning map of the competing candidates obtained directly from actual voting behavior. Second, as it will become clear later, when we discuss model estimation, our voter-choice model is estimated at a certain level of aggregation (precincts or districts), as actual voting data are seldom reported at the individual level. While we develop our model based on individual voter behavior, we show how this individual-level model translates into district-level voting and how it can be applied to data aggregated across a diverse population of voters. In doing so, we account for the unobserved heterogeneity in voter preferences within each district as well as across districts.

In modeling the actual voting behavior in an election, it is important to consider abstentions as they might represent a substantial proportion of the eligible votes. However, one must consider that the decision to cast a vote is related, but not the same as the decision to vote for a particular candidate. The choice probability given by equation (3) assumes that a vote is cast among the competing candidates. However, this choice decision is conditional on the fact that the voter decided to cast a vote. The decision to vote, on the other hand, depends on the maximum utility the voter expects from all available candidates (Train 2003), compared to the utility associated with the status quo (abstention):

$$P_{n(i)V} = \frac{\mathrm{e}^{\delta W_{n(i)}}}{\mathrm{e}^{V_{io}} + e^{\delta W_{n(i)}}}, \tag{4}$$

where

- $W_{n(i)} = \ln\left(\sum_{J=1}^{J} \mathrm{e}^{\alpha_j + \beta_j X_i + \lambda_j Z_{n(i)}}\right)$ is the "inclusive value" or expected maximum utility to voter n in district i provided by all available candidates,
- $V_{n(i)\mathrm{o}}$ is the utility associated with abstaining from voting; this utility is defined as for any real candidate in equation (2), except that the intercept α_o and demographic coefficients β_o are set to zero and the factor loadings are set to $\lambda_\mathrm{o} = -\sum_j \lambda_j$ for identification purposes; and
- δ is the coefficient of dissimilarity, which considers the possibility that candidates compete with each other in a different way they compete with abstentions. If $\delta = 1$ the model reverts to a multinomial logit model and abstentions are equivalent to a "null" candidate. If $\delta < 1$, political candidates compete more closely with each other than with abstentions, so that there is more cannibalization of votes among the candidates than with abstentions.

2.1 *Estimation of the Proposed Model on Aggregate (District or Precinct Level) Data*

The individual voter-choice model described in equations (1)–(4) may be estimated using individual choice intention or self-reported voting data commonly utilized in testing voter-choice models (Glasgow 2001). However, when applied to actual voting data, this type of model must be fitted to aggregate data because actual voting data are available only at a certain level of aggregation (precinct, district, or geographic region), due to ballot inviolability. For this purpose, we show below how our model can be estimated on aggregate (district or precinct levels) voting data.

Let y_{ij} be the number of votes cast at district i for candidate j, where $j = 0, 1, 2, \ldots, J$, including abstentions ($j = 0$). We specify the conditional likelihood of the voting data for district i as

$$p(y_i \mid Z_{n(i)}, \Theta) = \left[\frac{e^{\alpha_o+\beta_o X_i+\lambda_o Z_{n(i)}}}{e^{\alpha_o+\beta_o X_i+\lambda_o Z_{n(i)}} + e^{\delta \ln\left(\sum_{J=1}^{J} e^{\alpha_j+\beta_j X_i+\lambda_j Z_{n(i)}}\right)}}\right]^{y_{io}} \times \prod_{j=1}^{J}\left[\frac{e^{\ln\left(\sum_{J=1}^{J} e^{\alpha_j+\beta_j X_i+\lambda_j Z_{n(i)}}\right)}}{e^{\alpha_o+\beta_o X_i+\lambda_o Z_{n(i)}} + e^{\ln\left(\sum_{J=1}^{J} e^{\alpha_j+\beta_j X_i+\lambda_j Z_{n(i)}}\right)}}\right]^{y_{ij}}, \tag{5}$$

where $Z_{n(i)}$ is a p-dimensional vector of independent normal $\phi(\mu_i, \sigma)$ latent scores for voter n in district i and Θ collects all parameters of the model.

The unconditional likelihood of the voting data observed across all districts is obtained as the product of the conditional likelihood across all districts and by integrating out the independent normal latent scores $Z_{n(i)}$ across voters within each district and the independent standardized normal scores μ_i across districts. However, computing the likelihood function above involves double integration over two sets of p-dimensional standard multivariate normal densities, which is impractical for $p > 3$. Therefore, we estimate the model via simulated maximum likelihood (Gorieroux and Montfort 1997; Glasgow 2001; Train 2003), using a log-likelihood function that replaces integration with an average over T and S random draws:

$$\ell(\Theta \mid y_i) = \ln\sum_{s=1}^{S}\sum_{t=1}^{T}\left[\frac{e^{\alpha_o+\beta_o X_i+\lambda_o(Z_t+\sigma\mu_s)}}{e^{\alpha_o+\beta_o X_i+\lambda_o(Z_t+\sigma\mu_s)} + e^{\delta \ln \sum_{J=1}^{J} e^{\alpha_j+\beta_j X_i+\lambda_j(Z_t+\sigma\mu_s)}}}\right]^{y_{io}} \times \prod_{j=1}^{J}\left[\frac{e^{\ln \sum_{J=1}^{J} e^{\alpha_j+\beta_j X_i+\lambda_j(Z_t+\sigma\mu_s)}}}{e^{\alpha_o+\beta_o X_i+\lambda_o(Z_t+\sigma\mu_s)} + e^{\ln \sum_{J=1}^{J} e^{\alpha_j+\beta_j X_i+\lambda_j(Z_t+\sigma\mu_s)}}}\right]^{y_{ij}} - \ln(T^*S). \tag{6}$$

Once maximum-likelihood estimates $\hat{\Theta}$ are obtained, the posterior distribution of the factor scores for a district i can be obtained as

$$f(\mu_i \mid y_i, \hat{\Theta}) = \frac{\int p(y_i \mid Z, \hat{\Theta})\phi(Z \mid \mu, \hat{\sigma})dZ}{\int(\int p(y_i \mid Z, \hat{\Theta})\phi(Z \mid \mu, \hat{\sigma})dZ)\phi^*(\mu)d\mu}. \tag{7}$$

We obtain draws from this posterior distribution using the sampling-important resampling algorithm (see Wedel and Kamakura 2001 for details) and report the mean of this distribution as the district's score μ_i. Predictions of vote shares in each district are also obtained through simulations, to avoid the double integration over the distribution of latent scores within and across districts. Once the parameters of the model ($\hat{\Theta}$) and the factor scores ($\hat{\mu}_i$) for each district i are estimated using data from the first-stage election, the share of votes (s_{ij}) for remaining candidates j in the second stage in district i is predicted by

$$s_{ij}(\hat{\Theta}, \hat{\mu}_i) = \sum_{t=1}^{T}\left(\frac{e^{\hat{\alpha}_o+\hat{\beta}_o X_i+\hat{\lambda}_o(\hat{u}_i+\hat{\sigma}z_t)}}{e^{\hat{\alpha}_o+\hat{\beta}_o X_i+\hat{\lambda}_o(\hat{u}_i+\hat{\sigma}z_t)} + e^{\hat{\delta}\ln\sum_{J=1}^{J_2} e^{\hat{\alpha}_j+\hat{\beta}_j X_i+\hat{\lambda}_j(\hat{u}_i+\hat{\sigma}z_t)}}}\right) \times \left(\frac{e^{\ln\sum_{J=1}^{J_2} e^{\hat{\alpha}_j+\hat{\beta}_j X_i+\hat{\lambda}_j(\hat{u}_i+\hat{\sigma}z_t)}}}{e^{\hat{\alpha}_o+\hat{\beta}_o X_i+\hat{\lambda}_o(\hat{u}_i+\hat{\sigma}z_t)} + e^{\ln\sum_{J=1}^{J_2} e^{\hat{\alpha}_j+\hat{\beta}_j X_i+\hat{\lambda}_j(\hat{u}_i+\hat{\sigma}z_t)}}}\right), \tag{8}$$

where z_t is a p-dimensional vector of random draws from the standardized normal distribution.

2.2 *Implications of the Proposed Model to the Political Analyst*

The approach described in equations (1)–(8) provides the political analyst with a voter-choice model that is consistent with the theory of individual voter behavior, but is calibrated with actual voting data observed at some level of aggregation. Most importantly, the voter-choice model accounts for unobserved heterogeneity within and across districts and for observed (through the districts' demographic characteristics) heterogeneity across districts, thereby avoiding the assumption of proportional draw, which would ignore the possibility of cannibalization of votes among similar candidates, a phenomenon observed in two-stage elections, as we illustrate later.

The estimated intercepts α_j for each candidate j reflect the general strength of the candidate's political base across all voters, relative to abstentions, taken as a baseline. The demographic coefficients β_j measure the relative appeal (compared to abstentions) of candidate j to specific demographic constituencies (defined by the average demographic profile of each district). However, because the demographic predictors are only measured at the aggregate level (district), one must be careful in interpreting these coefficients, as these aggregate demographic predictors imply demographic homogeneity within each district (Achen and Shively 1995).

The latent-factor scores μ_i for a district i show the political inclinations for the average voter in district i in the p-dimensional latent space, relative to the average voter. Although one cannot relate directly these latent dimensions to specific issues or policies, these latent dimensions reflect how the districts differ in their preferences for the candidates and how well the candidates cater for the districts' preferences. The standard deviations (σ) reflect the heterogeneity in preferences within the districts; in order to avoid estimating one set of standard deviations for each district, we assumed that within-district heterogeneity is the same for all districts.

The factor loadings λ_j define the position of politician j in the latent issues or policies space; politicians positioned close to each other in the p-dimensional latent space appeal to the same districts and therefore cannibalize voter shares from each other. Factor models such as the one proposed here are known to be invariant to rotation (Wedel and Kamakura 2001), and therefore, interpretation of the latent dimensions themselves is highly subjective. However, this invariance to rotation allows us to directly interpret the relative position of the candidates defined by the factor loadings (λ) and preference weights for the districts, as we had done earlier.

The coefficient of dissimilarity δ indicates whether abstentions act as a null candidate or whether candidates compete more directly with each other for votes than with abstentions. A small value of δ suggests that the elimination of candidates is less likely to increase abstentions, and votes from the eliminated candidates are then likely to be taken by the remaining candidates in the second-stage election. When the value of this coefficient is close to 1, the original votes from the eliminated candidates will be distributed across all remaining alternatives, including abstentions and, therefore, elimination of a candidate in the second round will lead to a proportional increase in the number of abstentions.

3 A Predictive Test and Illustration of the Proposed Voter-Choice Model

To illustrate our proposed factor-analytic voter-choice model, we apply it to the voting data from the first stage of a two-stage election, and then use the estimates to predict the final outcome in the second stage. Two-stage elections are quite common throughout the world; they are used in at least 45 countries such as Argentina, Brazil, Chile, Dominican

Republic, Ecuador, France, Guinea-Bissau, Iran, Poland, Russia, Ukraine, etc. In this system, a large number of candidates compete in the first stage, and unless the winner obtains a minimum proportion of the valid votes, another election is run between the winner and runner-up, usually within 1–2 months of the first stage.

Two-stage elections provide an excellent opportunity for applying and testing voter-choice models. Politicians tend to allocate most of their campaign resources before the first stage as there is typically a relatively short time (usually less than a month) to campaign between the two stages. As many modern democracies utilize electronic voting systems, results from the first stage are often available within only a few days, and may therefore provide valuable information for the remaining candidates. Candidates who survive the first stage will want to know not only how well they did in relation to the other remaining contenders but also how voters for the eliminated candidates will shift their choices in the second stage. Usually, a large number of candidates compete in the first stage, and therefore it is possible that moderate candidates, who usually have more similar competitors than those in the extreme right or left, might cannibalize their votes in the first stage, whereas those in the extremes appeal to a more clearly defined niche. Another important question is whether voters who chose one of the eliminated candidates will bother to cast their ballots in the second stage. On one hand, these voters might be less motivated as their preferred candidate is not running; on the other hand, voters might become more motivated in the second stage, when the actual winner is chosen. By allowing for diversity in preferences not only across districts but also within each of the districts, our proposed model may potentially account for this cannibalization of votes among similar candidates. By nesting the choice of candidates under the decision to cast a ballot or not, our model also makes a clear distinction between the decision to vote and the choice of candidates.

The data we use in this illustration come from the 2002 state governor's election in the Brazilian state of Santa Catarina, for which we have election results for each of the 102 districts, each containing an average of 37,431 registered voters. We chose this election because it poses a challenging prediction problem. As shown in Table 1, six candidates competed in the first stage, with only two remaining in the second stage. However, the favorite candidate in the first stage, who had a clear advantage (32% more votes) over the runner-up in the first stage, lost in the second round by a small margin. Abstentions increased from the first to second stage, but not in proportion to the share of votes among the eliminated candidates, indicating that many of the voters who preferred the eliminated candidates shifted their choice to the remaining ones, rather than abstaining.

This pattern of results with nonproportional shifts of votes among the remaining options can only be accommodated with a model that avoids the proportional-draw property of the well-known multinomial logit model. A random-coefficients (mixed) logit model would account for some of these nonproportional draws only across districts. However, the data for the 102 districts show that shares have shifted nonproportionally also within each of the districts. In Fig. 1 we plot the actual ratio of votes between the two leading candidates in the two stages. A multinomial logit model, even one with random coefficients such as the mixed logit model, would allow for nonproportionality across all districts, but still predict a constant ratio for each of the districts (except for estimation errors), because it makes the assumption of *Independence of Irrelevant Alternatives* within each district. However, one can see clearly from Fig. 1 that the ratio of votes between the two leading candidates changed considerably in the second stage within each district, favoring the first-stage runner-up. This suggests that there is considerable cannibalization of votes among similar candidates within each district, which is understandable as there

Table 1 Voting data from Santa Catarina

	Party	*Votes*	*Share (%)*	*Ratio (PMDB)*
First stage	Abstentions	764,420	20.0	0.832
	PPB	1,217,059	31.9	1.325
	PMDB	918,615	24.1	
	PT	834,385	21.9	
	PSB	48,036	1.3	
	PPS	31,323	0.8	
	PSTU	4136	0.1	
Second stage	Abstentions	813,804	21.3	0.538
	PPB	1,491,723	39.1	0.986
	PMDB	1,512,447	39.6	

must be some diversity in political leanings across the 37,431 (on average) voters within each district.

Aside from accounting for the unobserved heterogeneity in voter preferences within and across districts, our proposed model also accounts for differences in preferences across districts that may be explained by observed characteristics of the precincts. For this purpose, we will use five socioeconomic characteristics, provided by electoral officials for the 102 districts, which we summarize in Table 2. However, we observed considerable multicolinearity (higher than 0.80) among the education-related predictors, and between these predictors and the income index (in the 0.50–0.70 range). This multicolinearity, along with the fact that these statistics represent averages across a large aggregation of voters, warrants caution in the interpretation of the response coefficients.

We applied different variants of our proposed model, along with a nested multinomial logit model to the data summarized in Tables 1 and 2. In order to avoid possible local optima in the simulated maximum-likelihood estimation, we ran the model multiple times from different random starts. Typically, one would select the appropriate number of latent dimensions (p) using information criteria such as the Bayesian information

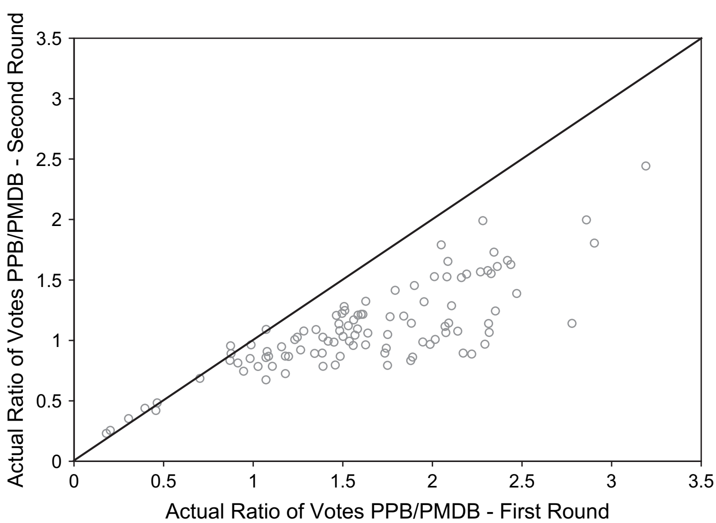

Fig. 1 Actual ratio of votes between the two leading candidates in the first and second stages.

Table 2 Summary statistics for the voting districts

Variable	*Average*	*Standard deviation*
Education index	0.904	0.037
Longevity index	0.814	0.034
Income index	0.733	0.054
School attendance	84.7	6.5
Literacy rate	93.2	3.1
Number of voters	37,431	20,073

criterion (BIC). However, because we are fitting our model to a large population, rather than a random sample, even these predictive criteria are likely to lead to overparameterization for predictive purposes; thus, they are not very informative. We chose the three-dimensional solution for ease of presentation, even though the BIC called for a four-dimensional solution.

Table 3 compares the fit of the different models both at estimation using data from the first-stage election and in predicting the results of the second stage. We compare the performance of the full formulation presented in equations (1)–(3), with several models that can be viewed as special cases of our model. The first special case (Model 2) only allows for unobserved heterogeneity in voter preferences across districts. The second variant (Model 3) allows for heterogeneous preferences within and across districts, but considers abstentions as another candidate. The third special case (Model 4) of the proposed model is similar to the previous one, except that it does not allow for heterogeneity within districts. The fourth contender (Model 5) is a nested logit choice model accounting only for observed voter heterogeneity across districts, through the demographics of each district.

Because all other models are restricted versions of our proposed model, it is not surprising that our model fits better to the calibration data (first-stage elections). Real predictive performance, however, is best measured in the prediction of the second round of elections, using the models estimated with first-round votes. The results displayed in Table 3 again demonstrate that our proposed model produced better predictions of the final election results, based on data from the first round. Once again, the two models accounting for within-district heterogeneity (Model 1 and Model 3) produced better predictions than the ones accounting only for heterogeneity across districts or no heterogeneity at all. The predictions from our full model have a root mean-squared error of 3.8 share points in predicting vote shares in the second stage, explaining 97.9% of the variance in vote shares across the three alternatives and 102 districts. We characterize this predictive performance as excellent, considering that these are true predictions out of sample and over time, and they do not incorporate any information about events that could have shifted voter preferences and perceptions between the two rounds of elections.

The main reason why the models accounting for heterogeneity in voter preferences within districts perform better than those models assuming homogeneity is that the former models allow for the possible cannibalization of votes among similar candidates within each district. This can be seen in Fig. 2, which plots the ratio of votes for the Partido Progressista Brasileiro (PPB) and Partido do Movimento Democratico Brasileiro (PMDB) candidates predicted by the full model and a simpler version assuming within-district homogeneity for the second-stage election, against the actual ratio observed in the first stage. Recall that the actual ratio between these two candidates drops considerably from the first to the second stage (see Fig. 1). Comparing the actual ratios in Fig. 1 with the

Table 3 Goodness-of-fit comparisons

	First-stage elections (fit)					*Second-stage elections (predictions)*			
Model	*Log likelihood*	*RMSE*	*1 − MSE/VAR (%)*	*RMSE (shares, %)*	*1 − MSE/VAR (shares, %)*	*RMSE*	*1 − MSE/VAR (%)*	*RMSE (shares, %)*	*1 − MSE/VAR (shares, %)*
1: Proposed (full)	−5,404,280	201	99.9	0.4	99.9	1479	96.9	3.8	97.9
2: Proposed with homogenous districts ($\sigma = 0$)	−5,408,280	433	99.5	1.2	99.3	2597	90.3	6.7	93.6
3: Proposed with abstentions as a candidate ($\delta = 1$)	−5,404,540	372	99.7	0.8	99.6	2026	94.1	5.2	96.2
4: Random-coefficients logit mapping ($\delta = 1$, $\sigma = 0$)	−5,409,000	434	99.5	1.2	99.2	2912	87.8	7.2	92.7
5: Nested logit ($\delta = 0$, $\lambda = 0$)	−5,532,430	2189	88.3	4.8	87.3	3596	81.4	8.0	91.5

MSE, mean-squared error; RMSE, root mean-squared error; VAR, variance.

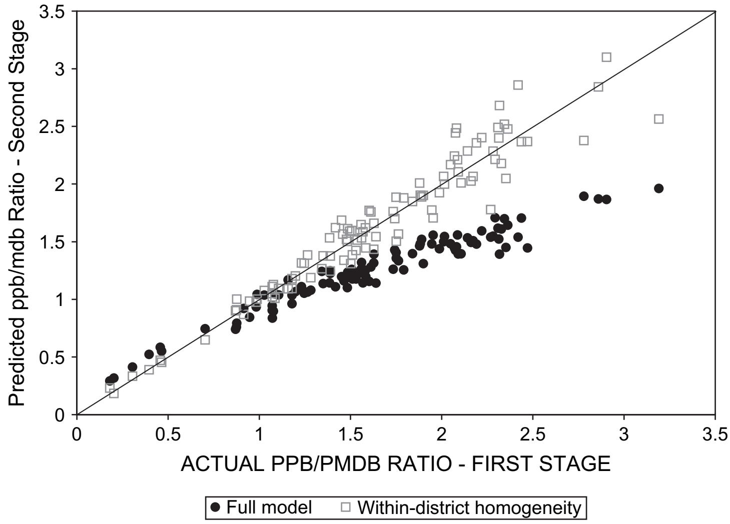

Fig. 2 Predicted ratio of votes between the two leading candidates in the second stage.

predicted ones in Fig. 2, one will note that the full model predicts a ratio for the second stage that better reflects the drop in the actual ratio observed between the first and second elections. The model assuming within-district homogeneity, on the other hand, implies independence of irrelevant alternatives within each district, thereby predicting the same ratio observed in the first stage.

The voting predictions for the second stage for each of the 102 districts obtained with the full model are compared in Fig. 3 with those obtained under the assumption of homogeneity in preferences within each district. Consistent with the performance measures reported in Table 3, Fig. 3 shows that the full model produced better predictions than the model ignoring heterogeneity within districts.

For the two candidates competing in the second round, the most critical prediction is the margin of votes between them. In Fig. 4, we compare the predicted margin of votes for the PPB over the PMDB party in the second round against the actual margin. The model tends to overpredict the margin for the PPB party, but correctly predicts the winner in most districts, except for those shown in the highlighted area, for which the margin of votes is relatively small.

We report the estimated parameters of the proposed model in Table 4, where one can see substantial within-district heterogeneity (σ) along the first two latent factors, considering that the district means have a standardized normal distribution. Table 4 also shows the problems anticipated with the multicolinearity among the education-related demographic predictors, as these predictors did not produce statistically significant effects on voting behavior. Based on the demographic coefficients (β) the PMDB party is more appealing than the PPB in districts with a higher longevity and income indices, ceteris paribus. The intercept estimates for each candidate suggest that the PMDB has a stronger general appeal than the runner-up, which is even dominated by the candidate in third place Partido dos Trabalhadores (PT) in terms of general appeal.

The estimate of the dissimilarity coefficient $\lambda = 0.861$ suggests that candidates compete with each other in a different way that they compete with abstentions. The estimates of the factor loadings are better understood in a chart (Fig. 5) mapping the position of the candidates in the latent issues space and their positions compared to the mapping of

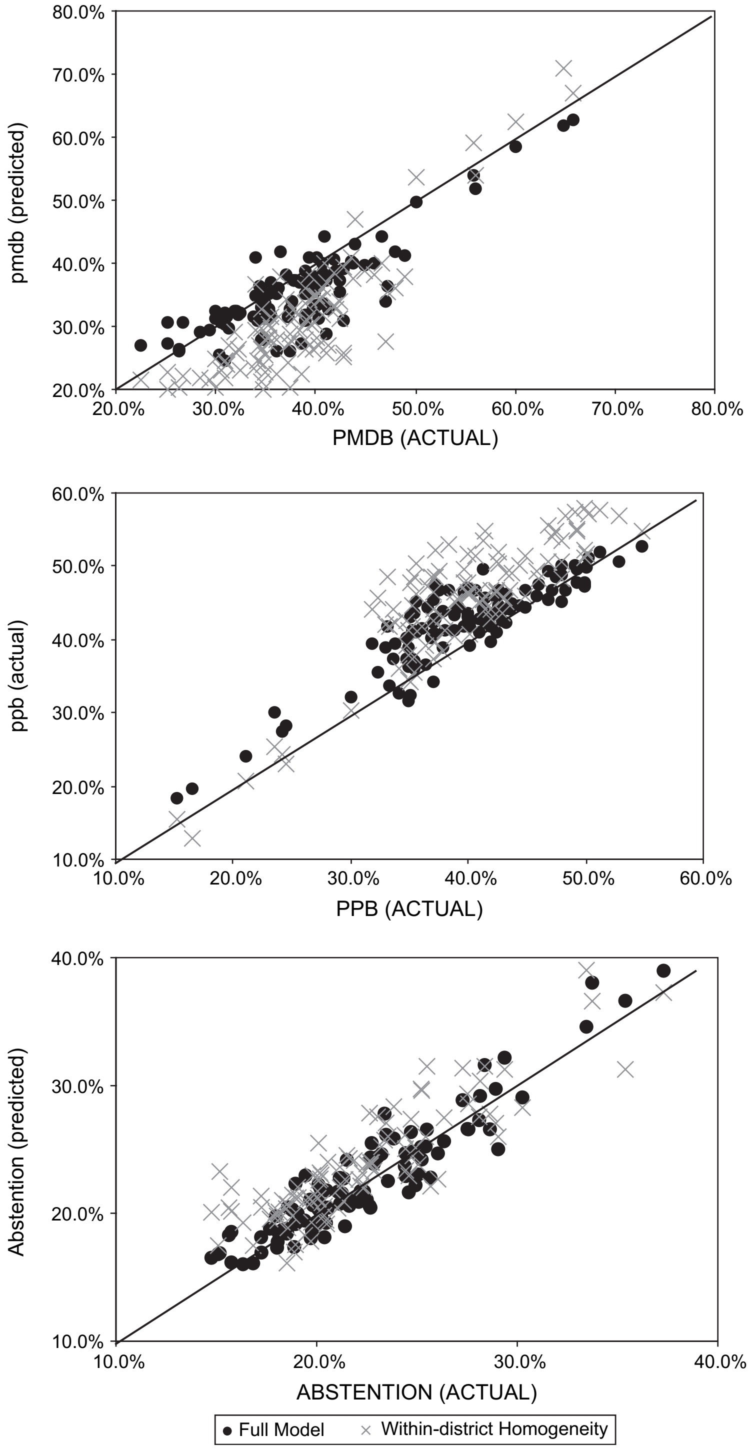

Fig. 3 Predicted versus actual shares of votes in the second-stage elections.

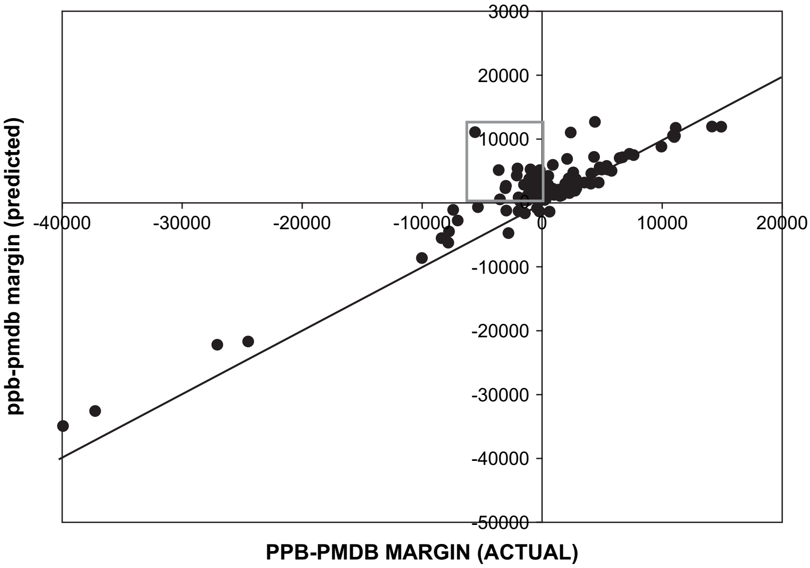

Fig. 4 Margin of votes for the PPB party across districts.

average political preferences within each district, shown in Fig. 6, where each vector shows the average direction of voter preferences within one particular district.

The mapping of the competing candidates in Fig. 5 suggests that the candidate from the PMDB party occupied a unique position relative to all the competing alternatives in the first stage. However, this unique position only accounts for the deviations in vote preferences in each district relative to the population average. The positioning of candidates in the first round may also provide some insights into potential alliances for the second round. Based on Fig. 5, one would expect the support by the Partido Popular Socialista (PPS) and Partido Socialista dos Trabalhadores Unificado (PSTU) in the second stage not to be of great value for the PPB candidate for two reasons: first, these two candidates did not attract many voters in the first round; second, according to Fig. 5, voters who preferred either of these two candidates are already likely to switch to the PPB candidate as the closest available alternative. A more interesting prospect for strategic alliances in the second round is the PT party, which was a strong third in the first round, and is fairly equidistant to both remaining candidates, suggesting that voters who preferred the PT party in the first round are likely to be swing voters in the second round. Even though the PT party is relatively close to abstentions in the positioning map, its followers are more likely to switch than to abstain because of the low coefficient of dissimilarity (δ), which suggests that candidates compete more closely with each other than with abstentions.

The preference vectors displayed in Fig. 6 show how the average preferences within each district deviate from the population of voters. Each vector represents one voting district and points to the direction in the latent issues space with the highest preference for the district, relative to the average voter. For example, Fig. 6 shows that the vector for district A points in the direction where PPB is located in Fig. 5; this suggests that the typical voter in district A is more likely than the average voter in the population to vote for the PPB candidate, while the typical voter in district B is more likely than the average voter in the population to vote for the PMDB candidate. Looking at these preference vectors, the analyst can easily predict the voting inclinations for each district in the second stage.

Table 4 Parameter estimates for the full proposed model

		Factor loadings (x)			*Demographic coefficients (β)*				
Candidate	*Intercept (γ)*	*Factor 1*	*Factor 2*	*Factor 3*	*Education*	*Longevity*	*Income*	*Schooling*	*Literacy*
ABST	0.000	0.094*	0.197*	−0.161*	0.000	0.000	0.000	0.000	0.000
PPB	0.719*	0.262*	0.012	0.255*	0.257	−0.035*	−0.273*	0.089	−0.022
PMDB	0.024	−0.325*	0.843*	0.003*	−0.357	0.133*	0.328*	−0.109	0.596*
PT	0.321*	−0.194*	−0.391*	−0.162*	0.184	0.076*	−0.130*	−0.192	0.034
PSB	−2.669*	−0.077*	−0.097*	−0.253*	−0.447	0.015*	0.111*	−0.073	0.669
PPS	−3.175*	0.195*	−0.336*	0.263*	0.056	−0.078*	0.331*	0.105	−0.160
PSTU	−5.141*	0.045*	−0.228*	0.055*	0.170	0.014	0.128*	−0.033	0.079
Within-precinct standard deviation (σ)		1.695*	1.922*	0.067*					
Dissimilarity coefficient (λ)	0.861*								

Note. *Statistically significant at the 0.05 level.

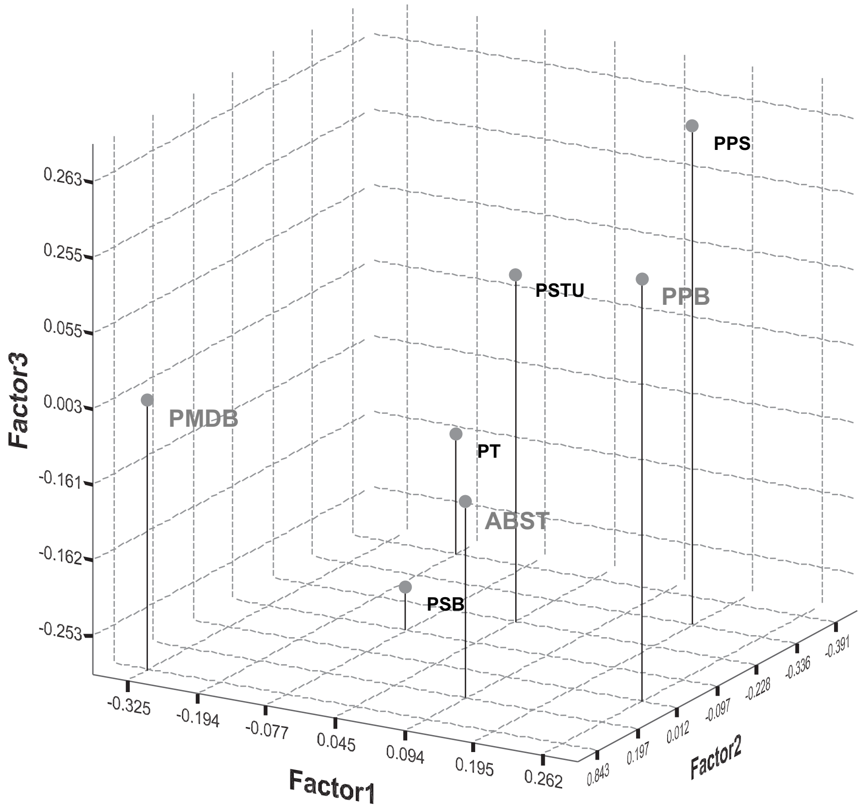

Fig. 5 Latent space of political candidates in the first-stage election.

4 Conclusions and Directions for Future Research

There are at least three reasons why the modeling and prediction of multiparty elections is increasingly important. First, as democracy spreads throughout the world, we see more electoral systems with a multitude of competing parties covering a broad political spectrum. Second, two-stage elections, where a large number of candidates are reduced to a few (usually two) in a first stage and then compete in a second election within a few weeks from the first, are becoming the norm rather than the exception, making it possible to calibrate a model on the results from the first round and test it in making real predictions for the second round. Third, most modern democracies now collect and tally votes electronically, making results available at the precinct or district levels within a couple of days from the election, which makes it possible for political analysts to calibrate their models on the results from the first round, thus providing politicians with valuable diagnostic information to guide their strategy for the short campaign before the second round of elections.

The factor-analytic nested logit model we proposed in this study offers several features that make it suitable for the analysis of the actual voting data obtained from multiparty elections, especially those happening in two stages. By allowing voters to differ in their political preferences not only across districts but also within each district, our model accounts for the inevitable shifts in voting behavior after many of the candidates are eliminated in the first round of two-stage elections. We provide hard evidence that, at

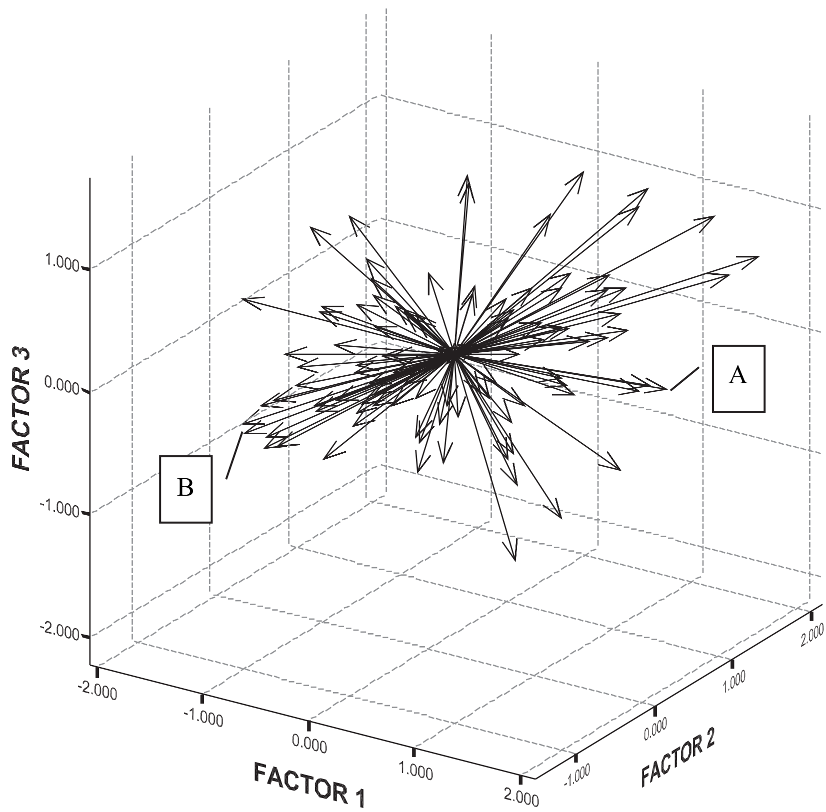

Fig. 6 Vectors indicating average political preferences in each district.

least in the election we studied, these shifts do not follow the proportional draws predicted by the popular multinomial logit model at the aggregate level, or even by the more sophisticated mixed logit model at the district level. Our proposed model avoids this proportional-draw assumption at the aggregate and at the district levels.

Another distinguishing feature of our proposed model, relative to previous voter-choice models and models of aggregate vote shares, is that instead of using self-reported perceptual measures of the political candidates, which must be defined a priori and measured on a sample of potential voters, we infer the relative position of the competing candidates on a latent issues space directly from actual voting data. This latent space shows how districts differ from the average voter in their political preferences and how the candidates are positioned in relation to these preferences and to their competitors. Political strategists may use this latent space to understand their candidate's position relative to voter preferences and to the competition. In a two-stage election, the map obtained from first-round votes may be useful for the surviving candidates in identifying weaknesses and strengths at the district level and in defining strategic alliances with eliminated candidates.

Obviously, predictions for the second stage based on the model calibrated at the first stage are conditional on the assumption that preferences and perceptions remain stable, which is only likely to hold within a short time period and in the absence of dramatic events affecting political preferences and perceptions between the two stages. Therefore, the model is more useful as a tool for understanding voter preferences and the competitive standing of the candidates in the first round than in making predictions for the second round. Predictions for the second round should be viewed as conditional forecasts, under the scenario that nothing is done to change the political landscape between the two rounds of elections.

One possible avenue for future extensions of the proposed model would be to incorporate information about events that might affect the competitive positioning of the remaining candidates, such as political alliances and targeted campaigning in "swing" districts. Such an extension would not only improve forecasting for the second round but could also be used to consider different scenarios regarding the campaign strategy by the remaining candidates between the two rounds.

References

Achen, Christopher, and W. Phillips Shively. 1995. *Cross level inference.* Chicago, IL: University of Chicago Press.

Black, D. 1958. *The theory of committees and elections.* New York: Cambridge University Press.

Cho, S., and J. W. Endersby. 2003. Issues, the spatial theory of voting, and British general elections: A comparison of proximity and directional models. *Public Choice* 114:275–93.

Coughlin, P. J. 1992. *Probabilistic voting theory.* Cambridge: Cambridge University Press.

Coughlin, P. J., and Nitzan, S. 1981. Electoral outcomes with probabilistic voting and Nash social welfare maxima. *Journal of Public Economics* 15:113–22.

Dow, J. K., and J. W. Endersby. 2004. Multinomial probit and multinomial logit: A comparison of choice models for voting research. *Electoral Studies* 23:107–22.

Downs, A. 1957. *An economic theory of democracy.* New York: Harper & Row.

Elrod, T. 1988. Choice map: Inferring a product-market map from panel data. *Marketing Science* 7:21–40.

Enelow, J. M., and M. J. Hinich. 1984. *The spatial theory of voting.* New York: Cambridge University Press.

Glasgow, Garret. 2001. Mixed logit models for multiparty elections. *Political Analysis* 9(1):116–36.

Gourieroux, C., and A. Monfort. 1997. Simulation based econometric methods. Oxford: Oxford University Press.

Honacker, James, Jonathan N. Katz, and Gary King. 2002. A fast, easy and efficient estimator for multiparty electoral data. *Political Analysis* 10(1):84–100.

Jackson, John E. 2002. A seemingly unrelated regression model for analyzing multiparty elections. *Political Analysis* 10(1):49–65.

Katz, Jonathan N., and Gary King. 1999. A statistical model for multiparty electoral data. *American Political Science Review* 93(1):15–32.

MacDonald, S. E., G. Rabinowitz, and O. Listhaug. 2001. Sophistry versus science: On further efforts to rehabilitate the proximity model. *Journal of Politics* 63:482–500.

McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontier in Econometrics*, ed. P. Zarembka, 105–42. New York: Academic Press.

Mikhailov, Nikolai, Richard G. Niemi, and David L. Weimer. 2002. Application of Theil group logit methods to district level vote shares: Tests of prospective and retrospective voting in the 1991, 1993, and 1997 Polish elections. *Electoral Studies* 21:631–48.

Quinn, Kevin M. 2004. Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis* 12(4):338–53.

Tomz, Michael, Joshua A. Tucker, and Jason Wittenberg. 2002. An easy and accurate regression model for multiparty electoral data. *Political Analysis* 10:66–83.

Train, Kenneth E. 2003. *Discrete choice methods with simulation.* Cambridge, UK: Cambridge Press.

Zeng, L. 2000. A heteroscedastic generalized extreme value discrete choice model. *Sociological, Methods & Research* 29(1):119–45.

Wedel, Michel, and Wagner A. Kamakura. 2001. Factor analysis with mixed observed and latent variables in the exponential family. *Psychometrika* 66(4):515–30.