Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 22 (2006) 689–706
www.elsevier.com/locate/ijforecast

Modeling voter choice to predict the final outcome of two-stage elections

Wagner A. Kamakura [a,*], José Afonso Mazzon [b], Arnaud De Bruyn [c]

[a] Fuqua Graduate School of Business, Duke University, 1 Towerview Rd. Durham, NC 27708, United States
[b] Faculdade de Administração e Economia, University of São Paulo, Ave. Prof. Luciano Gualberto, 908 CEP 0558-900 São Paulo, Brazil
[c] ESSEC Business School, Avenue Bernard Hirsch, 95000 Cergy, France

Abstract

Most election forecasting research to date has been conducted in the context of single-round elections. However, more than 40 countries in the world employ a two-stage process, where actual voting data are available between the first and the second rounds to help politicians understand their position in relation to each other and to voter preferences and to help them predict the final outcome of the election. In this study we take advantage of the theoretical foundation on voter behavior from the political science literature and the recent methodological advances in choice modeling to develop a Nested Logit Factor Model of voter choice which we use to predict the final outcome of two stage elections and gain insights about the underlying political landscape. We apply the proposed model to data from the first stage and predict the final outcome of two stage elections based on the inferences made from the first stage results. We demonstrate how our proposed model can help politicians understand their competitive position immediately after the first round of actual voting and test its predictive accuracy in the run-off election across 11 different state governorship elections.

Keywords: Voter choice; Election forecasting; Political marketing; Nested Logit model

1. Introduction

The main purpose of this study is to propose a voter choice model with two main features of value to political analysts and candidates competing in two-stage elections. First, when calibrated on the results from the first round of elections, our model provides the political analyst with a positioning map for all competing candidates jointly with a preference map depicting voter preferences in each precinct. Second, the model allows politicians to predict, based on the actual results from the first stage, the final outcome of two-stage elections at the precinct level.

Since the seminal work of Bean (1948), followed by a renewed contemporary interest triggered by the work of Kramer (1971) and Tufte (1978), forecasting elections has been a major research topic in the

* Corresponding author. Tel.: +1 919 660 7855.
E-mail addresses: kamakura@duke.edu (W.A. Kamakura), jamazzon@usp.br (J. Afonso Mazzon), debruyn@essec.fr (A. De Bruyn).

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2006.04.002

political science literature. Understanding voter preferences and forecasting the final outcome of elections is of critical importance to politicians, as they can use the insights gained from the exercise to fine-tune their campaign strategies. For this reason, one already finds an extensive literature on the prediction of election outcomes. For example, the *International Journal of Forecasting* devoted a special issue to the topic (1999, volume 15, issue 2). Similarly, the *American Politics Quarterly* published a collection of essays on forecasting the 1996 US presidential election (October 1996).

Most methods used in practice base their forecasts on opinion polls, mainly because this source of information is abundant and readily available before elections. However, academics have long recognized that opinion polls are not without limitations as measurements of political perceptions and preferences and hence as predictors of actual voting behavior. First, since polls are often based on relatively small samples, they might not reflect the true diversity in a highly heterogeneous population ranging from extreme left to extreme right political orientations. Second, simulated voting in "trial heat" polls represent *stated* rather than *revealed* voter preferences. Consequently, voters might (1) exhibit strategic behavior and try to send signals to the candidates in their reported voting intention, (2) form their true preferences only when faced with the real decision at the voting booth, or (3) not be truthful to pollsters because they are embarrassed to reveal their preferences for certain candidates, especially if surveys are conducted face-to-face (see Jérôme, Jérôme, & Lewis-Beck, 1999, p. 167). Finally, as argued by Gelman and King (1993), polls reflect short-term responses to daily campaign events which might have less impact in the eventual outcome of the election.

Recent debacles of opinion polls in predicting election outcomes in the US, in Britain (Rallings & Thrasher, 1999), and in France (Dasgupta & Maskin, 2004; Jérôme et al., 1999) have highlighted the limitations of such sample-based, stated-preference methods. Along the same lines as Armstrong's work, which showed that averaging across methods could reduce forecasting errors (Armstrong, 2001), academics have suggested methods to improve the predictive accuracy of preference polls, such as pooling different information sources (Polly, 2005), weighting polls based on how many months before the elections they were conducted (Holbrook & DeSart, 1999), or including votes from previous elections to supplement poll data (Brown & Chappell, 1999).

A refinement of opinion polls is to use voters as forecasters. Lewis-Beck and Tien (1999) show that micromodels derived from *vote expectations* turn out to be more accurate than those derived from *vote intentions*; that is, asking the question in terms of how *others* will vote seem to produce more reliable estimates than when it is framed in terms of how *oneself* will vote. Electronic markets are a promising extension of this approach which use bidding mechanisms to elicit precise marketplace expectations of election results. In particular, the Iowa Electronic Markets have gathered extensive data about the accuracy of such methods for political markets (Berg, Forsythe, Nelson, & Rietz, 2000).

Macromodels based on economic and political fluctuations have also been offered as alternative solutions. Despite some criticisms (Greene, 1993), these models have become predominant in the literature, using statistical modeling from systematic observation of indicators such as employment rate, health of the economy or political stability to predict election outcomes (e.g., Brown & Chappell, 1999; Jérôme et al., 1999; Stambough & Thorson, 1999). The foundation of this approach lies in the implicit assumption that election results are based primarily on the (economic) performance of the party controlling the country and that a positive track record should lead to the victory of the incumbent political party.

A review of the election forecasting literature shows that, "with few exceptions, forecasting studies have focused on the [US] presidential outcome" (Holbrook & DeSart, 1999, p. 137), while most of the so-called "exceptions" focus on statewide US elections. This narrow focus on single-round (US-style) elections has led academics to concentrate on predictions based on sample surveys where only attitudinal data are available and to overlook two-stage elections—a type of election used in more than 40 countries in the world as shown in Table 1.

Two-stage elections offer politicians valuable information about voter preferences at the precinct level before the final voting, since actual behavioral data are available after the first round. In these election systems, typically a large number of candi-

Table 1
Countries with elections in two turns by continent

	([a])
America (11)	
Argentina	28
Bolivia	–
Brazil	21
Chile	35
Costa Rica	63
Dominican Republic	–
Ecuador	34
El Salvador	–
Guatemala	45
Nicaragua	–
Uruguay	30
Europe (13)	
Austria	–
Bulgaria	7
Croatia	14
Finland	21
France	14
Lithuania	14
Macedonia	14
Poland	14
Portugal	–
Romania	14
Slovakia	14
Slovenia	14
Ukraine	56
Africa (15)	
Benin	14
Burkina Faso	–
Central African Republic	56
Chad	39
Comoros	–
Congo, Republic of	28
Ghana	21
Guinea Bissau	49
Madagascar	–
Mali	14
Mozambique	–
Niger	21
Sao Tome and Prince	–
Senegal	21
Sudan	–
Asia (5)	
Afghanistan	–
Armenia	14
Egypt	10
Iran	7
Russia	30

[a] Typical number of days between the first and the second turn of the elections, when available.

dates compete in the first stage. Unless the candidate receiving the most votes reaches a minimum share of valid votes, a second stage election, in which a subset of the candidates compete, is held within a few weeks from the first election.

Two-stage elections also represent an invaluable opportunity to apply theoretical and methodological advances in the modeling of voter choice. First, the results from the first stage, usually available within a few days of the election at the district or precinct level, represent voters' revealed preferences rather than stated preferences (as commonly used in poll-based models of voter choice). Second, due to the typically short time period between the two rounds of the elections (*median* = 21 days, *mean* = 25 days, see Table 1), voters are less likely to change their perceptions of the remaining candidates than during the long campaign before the first round and even less likely to change their political preferences or tendencies after the first round, unless dramatic events produce these shifts. Third, the insights about voter perceptions and preferences obtained from the first stage can be of value not only for candidates retained for the second stage, but also for those eliminated. Both winners and losers in the first stage may use these insights about voter preferences to form strategic alliances for the second round, as we show later. Fourth, revealed patterns of voter preferences across precincts provide valuable insights on where each candidate is vulnerable and where she might have a better chance of gaining more votes, allowing politicians to fine-tune their campaign at the local level. Finally, analysis of the voting data across precincts may produce insights about the demographic profile of the political bases for each candidate in the first round, which can help remaining candidates in the second round better understand their strengths/weaknesses across demographic segments. Therefore, the value of a forecasting model in a two-stage election goes beyond merely predicting the final outcome, but should also lie in providing a roadmap for political candidates to attempt to modify the predicted outcome. A model that produces these forecasts based on insights about voter perceptions and preferences may help political analysts and candidates fine-tune their campaign strategy in the few weeks remaining before the final round of elections to: (a) strengthen their position, (b) better understand the competitive structure of the electoral market and how they are positioned compared to the other candidates, (c) identify their real strengths and weaknesses, and (d) perhaps attempt to avoid the predicted outcome.

Predicting the final outcome of two-stage elections based on the results from the first stage poses several challenges to the analyst. First, data is available only at a certain level of aggregation such as electoral zones, districts or precincts. While some advances have been made in inferring individual-level voting behavior from aggregate data (King, 1997), some debate still exists on the robustness of these ecological inferences to violations of the underlying assumptions of the model (Cho, 1998; Freedman, Klein, Ostland, & Roberts, 1998). Therefore, one must decide whether to model behavior at the observed level of aggregation or to make ecological inferences of individual-level behavior.

Second, a well-known problem with first-choice voting systems used in most elections is the cannibalization of votes among similar candidates, leaving moderate candidates in the center competing against several other moderates while extremists face less direct competition, as they focus on smaller fringe-segments. One example of this phenomenon was the 2002 French presidential election, where as expected, Jacques Chirac, the right-winger incumbent received the most votes in the first stage (19.9%). The great surprise was the extreme right-winger Le Pen's second place with 16.9% of the votes. Jospin, another favorite for the run-off along with Chirac was eliminated with 16.2% of the votes (Dasgupta & Maskin, 2004) to the astonishment of forecasters who consistently predicted Jospin to enter in the second round based on public opinion polls. As candidates are eliminated in the first round, the model for predicting the second-round results must take into account the fact that the political appeal of the surviving candidates might have been "diluted" in the previous round due to cannibalization among similar candidates, as demonstrated in the 2002 French presidential election.

Third, voter involvement in the elections might vary between the first and second rounds. On one hand, voters might be more engaged in the first round when they can support a candidate (among many) that truly represents their values. On the other hand, they might become more engaged in the second round as this election is of immediate consequence to the final outcome.

Fourth, there are not easily measurable, stable and widely accepted features or attributes for political candidates that can capture all the nuances of the candidates' traits or positions on political issues and that can be used as exogenous variables to explain voter choices. Perceptual measures of the candidates are obviously only available for samples of potential voters (Cho & Endersby, 2003) and do not necessarily represent the views of all voters or reflect the differences in perceptions across precincts. Moreover, Klein and Ahluwalia (2005) have shown that voters' assessments of political candidates' traits or positions were biased by respondents' political preferences (negativity bias). The presence of strong confounding between perceptual measures of the candidates and voters' preferences for the candidates hence make the first measure unfit to reliably predict the latter.

The voter-choice model we propose attempts to overcome some of these challenges by taking advantage of the voter preferences revealed in the first round of elections in order to provide politicians with a depiction of the competitive structure in the first round and predictions of the results for the second round. Our proposed model:

- Accounts for *observed* heterogeneity in voter preferences across precincts, based on the demographic profile of these precincts, when available. In doing so, the model provides insights into the appeal of each candidate to different demographic groups.
- Accounts for *unobserved* heterogeneity in voter preferences across precincts, thereby considering the cannibalization of votes among similar candidates in the first round.
- Estimates a latent factor structure that provides the relative position of each candidate in a latent space, along with directional measures of voter preferences for each precinct in the same latent space. This latent factor structure can be interpreted as the competitive mapping of candidates on a latent preference space for the precincts, providing candidates with a summary depiction of their standing relative to competitors and relative to the preferences across precincts.
- Uses a nested-Logit formulation to account for the fact that the voter decision to abstain in the election

is different from the choice among the available candidates, thereby providing a better accounting for abstentions than the popular multinomial Logit formulation commonly used in modeling voter choice.

2. Modeling voting behavior and inferring voter preferences

As mentioned earlier, our main goal is to model voting behavior in the first round of two-stage elections to obtain estimates of voter perceptions of the candidates and of their preferences for these candidates. Our model is grounded on the classic spatial theory of voting (Black, 1958; Downs, 1957). According to this theory, voters make their choices through a comparison of their own preferences on policies and their perceptions of the candidates' positions on those issues. We do not account for the possibility that voters might use their first-stage choices strategically, and assume they will cast their ballots on the candidate they perceive as best for them. Classical "Downsian" theory specifies an 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 & Hinich, 1984). More recently, political scientists have used a directional theory (MacDonald, Rabinowitz, & Listhaug, 2001), where voters reward candidates who take a strong and clear position on issues and propose to move policies in direction preferred by voters (Cho & Endersby, 2003). We follow the later directional or vector perspective in the model we propose below.

However, contrary to the political science literature, which uses self-reported perceptual measures of the candidates on pre-defined "issues" or "policies" to explain voting intentions, we will infer a latent "issues" space, defined by the position of each candidate and the directional preferences for each precinct, directly from the voting behavior observed in the first stage of a two-stage election. We do so because perceptual measures of the candidates are rarely available at the precinct level and are unlikely to capture all the issues and policies that differentiate the candidates. For this purpose we adopt a latent-variables formulation originally proposed by Elrod (1988). With this latent factor approach, we let the actual voting behavior determine the dimensionality and nature of the latent space differentiating the politicians relative to voter preferences.

Following the probabilistic voting literature (Coughlin, 1992; Coughlin & Nitzan, 1981; Zeng, 2000), we assume that many unobservable factors beyond the candidates' position along the voters' preferred directions and observed characteristics of the voter may affect voter behavior, leading to a stochastic model (Dow & Endersby, 2004; Paap et al., 2005; Zeng, 2000). We assume that the value of candidate j to voters in precinct i depends on observable and unobservable characteristics of the precinct and the candidate, as well as a random component that captures the effects of all other factors not explicitly considered in the model. We formulate the total value of the candidate j to precinct i as

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \alpha_j + \beta_j X_i + \lambda_j Z_i + \varepsilon_{ij} \tag{1}$$

where

- α_j is a candidate-specific intercept, representing the general appeal (or political equity) of candidate j across all voters.
- X_i is a k-dimensional vector with the known demographic profile of precinct i.
- β_j is a k-dimensional vector of demographic coefficients representing candidate j's appeal to a particular demographic constituency, relative to other candidates.
- Z_i is a p-dimensional vector of latent scores (to be estimated) capturing unobserved deviations in voter preferences for precinct i, relative to the population average, assumed to be independent, identically distributed standardized normals, without loss of generality.
- λ_j is a p-dimensional vector of factor weights (or loadings) for candidate j, representing the candidate's location in the latent "issue" space of voter heterogeneity.
- ε_{ij} are random components of utility, assumed to be i.i.d. extreme-value variables across candidates and precincts.

Note that even though we assume the random errors ε_{ij} to be independent, the random values U_{ij} are correlated across precincts and candidates; this correlation is captured by the factor structure ($\lambda_j Z_i$), as shown by Wedel and Kamakura (2001). This factor structure accounts for the unobserved diversity in voter preferences across precincts. By allowing for unobserved heterogeneity in preferences across precincts, the proposed model will avoid the proportional draw assumption implicit in the popular multinomial Logit model, thereby accounting for the cannibalization of votes among similar candidates in the first-stage election. Accounting for cannibalization among similar candidates is essential in modeling the choice behavior in two-stage elections because typically a large number of candidates in the first stage are 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, as we described earlier for the 2002 French presidential elections.

Observed differences in preferences are accounted through the demographic profiles of the precincts. These observed and unobserved differences in preferences across voters will explain cannibalization of votes among similar candidates (Kamakura & Russell, 1989).

In developing this model, we also want to take into account the fact that the decision to cast a vote is related to, but not the same as, the decision to vote for a particular candidate. In other words, when making the voting decision, the voter evaluates all available candidates and decides whether it is worth casting a vote. Therefore, the choice of candidate is nested under the voting decision. The probability that voter in precinct i will vote on candidate j, *conditional on her casting a vote*, is given by the well-known Multinomial Logit model, widely used in modeling voter choice behavior (Dow & Endersby, 2004; Paap et al., 2005):

$$P_{ij/\mathrm{V}} = \frac{\mathrm{e}^{\alpha_j+\beta_j X_i+\lambda_j Z_i}}{\sum_{j'} \mathrm{e}^{\alpha_{j'}+\beta_{j'} X_i+\lambda_{j'} Z_i}}. \tag{2}$$

The choice decision above 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 candidate expects from all available candidates (Train, 2003), compared to the utility associated with abstention:

$$P_{\mathrm{V}} = \frac{\mathrm{e}^{\delta W_i}}{\mathrm{e}^{V_{i0}} + \mathrm{e}^{\delta W_i}}, \tag{3}$$

where

- $W_i = \ln\left(\sum_{J=1}^{J} \mathrm{e}^{\alpha_j+\beta_j X_i+\lambda_j Z_i}\right)$ is the "inclusive value" or expected maximum utility to precinct i provided by all available candidates.
- V_{i0} is the utility associated with abstaining from voting. This utility is defined as for any real candidate in (1), except that the intercept α_0 and demographic coefficients β_0 are set to zero, and the factor loadings are set to $\lambda_0 = -\sum_j \lambda_j$ for identification purposes.
- δ is the coefficient of dissimilarity, which defines the interdependence between the political candidates and abstention. If $\delta = 1$ the model reverts to a multinomial Logit model. 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. The coefficient δ also defines the importance of voting (relative to abstention) among all voters.

2.1. Model estimation

At the very best, voting data is only available for each ballot box, and in most cases only at the precinct or electoral-zone level, due to ballot inviolability. While one might think of inferring the distribution of individual voter preferences based on these aggregate data using some form of ecological inference (King, 1997), our main purpose is to use the estimated model to make predictions for the second stage, for which we only have data on the exogenous variables at the same level of aggregation (precincts). Therefore, we infer unobserved heterogeneity in voter preferences only across precincts.

Let y_{ij} be the number of votes cast in precinct 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 precinct i as:

$$f(y_i|Z_i,\Theta) = \left[\frac{\mathrm{e}^{\alpha_0+\beta_0 X_i+\lambda_0 Z_i}}{\mathrm{e}^{\alpha_0+\beta_0 X_i+\lambda_0 Z_i} + \mathrm{e}^{\delta \ln\left(\sum_{J=1}^{J} \mathrm{e}^{\alpha_j+\beta_j X_i+\lambda_j Z_i}\right)}}\right]^{y_{i0}} \times \prod_{j=1}^{J}\left[\frac{\mathrm{e}^{\ln\left(\sum_{J=1}^{J} \mathrm{e}^{\alpha_j+\beta_j X_i+\lambda_j Z_i}\right)}}{\mathrm{e}^{\alpha_0+\beta_0 X_i+\lambda_0 Z_i} + \mathrm{e}^{\ln\left(\sum_{J=1}^{J} \mathrm{e}^{\alpha_j+\beta_j X_i+\lambda_j Z_i}\right)}}\right]^{y_{ij}} \tag{4}$$

where Z_i is a p-dimensional vector of i.i.d. standardized normal latent scores for precinct i and Θ collects all parameters of the model.

The unconditional likelihood of the voting data observed across all precincts is obtained as the product of the conditional likelihood across all precincts and by integrating out the i.i.d. standardized normal latent scores Z:

$$L(\Theta|Y) = \prod_i \int f(y_i|Z_i,\Theta)\phi^*(Z_i)\mathrm{d}Z_i. \quad (5)$$

Computing the likelihood function above involves integration over a p-dimensional standard multivariate normal density, which can be cumbersome for $p>3$. Therefore, we estimate the model using simulated maximum likelihood (see Gourieroux & Montfort, 1997 for details), using a log-likelihood function that replaces integration with an average over T random draws:

$$\ell(\Theta|y) = \sum_i \left(\ln \sum_{t=1}^{T} \left[\frac{\mathrm{e}^{\alpha_0+\beta_0 X_i+\lambda_0 Z_{it}}}{\mathrm{e}^{\alpha_0+\beta_0 X_i+\lambda_0 Z_{it}} + \mathrm{e}^{\delta \ln\left(\sum_{J=1}^{J} \mathrm{e}^{\alpha_j+\beta_j X_i+\lambda_j Z_{it}}\right)}} \right]^{y_{i0}} \times \prod_{j=1}^{J} \left[\frac{\mathrm{e}^{\ln\left(\sum_{J=1}^{J} \mathrm{e}^{\alpha_j+\beta_j X_i+\lambda_j Z_{it}}\right)}}{\mathrm{e}^{\alpha_0+\beta_0 X_i+\lambda_0 Z_{it}} + \mathrm{e}^{\ln\left(\sum_{J=1}^{J} \mathrm{e}^{\alpha_j+\beta_j X_i+\lambda_j Z_{it}}\right)}} \right]^{y_{ij}} - \ln T \right). \quad (6)$$

Estimates of the parameters in Θ are obtained as the values that maximize the likelihood function above. Technical details about the estimation of choice models via simulation can be found in Gourieroux and Montfort (1997) and Train (2003). Once maximum-likelihood estimates $\hat{\Theta}$ are obtained, the posterior distribution of the factor scores for a precinct i can be obtained as,

$$p(Z_i|y_i,\hat{\Theta}) = \frac{f(Z|y_i,\hat{\Theta})}{\int f(Z|y_i,\hat{\Theta})\phi^*(Z)\mathrm{d}Z}. \quad (7)$$

We obtain draws from this posterior distribution using the sampling-important re-sampling algorithm and report the mean of this distribution as the precinct's score Z_i. Technical details can be found in Wedel and Kamakura (2001).

2.2. Implications of the proposed model

Estimation of the proposed Nested Logit Factor Model on the voting results from the first round of elections will provide politicians with a summary of voter preferences revealed by their voting behavior across all precincts. The estimated intercepts α_j for each candidate j provide a measure of the candidate's general political "equity" or "capital" across all precincts, relative to abstentions. This estimate reflects the general strength of the candidate's political base regardless of the demographic composition and can be used by a candidate eliminated in the first round to bargain for political benefits in exchange for her support of another candidate competing in the second round.

The demographic coefficients β_j measure the relative appeal (compared to abstentions) of candidate j to specific demographic constituencies (defined by the demographic profile of each precinct). For example, candidate A can compare himself with candidate B on a demographic characteristic k by looking at $\mathrm{e}^{(\beta_{\mathrm{A}k}-\beta_{\mathrm{B}k})}-1$, which measures the percent increase in the odds of voting for candidate A relative to B, for each unit of the demographic variable k in a precinct. This comparison may be valuable to the remaining candidates in the second stage, providing them with an indication of their appeal to certain demographic groups relative to the other remaining candidate.

The latent factor scores z_i for a precinct i show the directions of maximum voting preferences in the p-dimensional latent space. In other words, voters in this precinct are more likely than others to vote on candidates positioned far from the origin along the direction defined by these scores. While one may not relate these latent dimensions directly to specific issues or policies, these latent dimensions reflect how the precincts differ in their preferences for the candidates. Therefore, as long as each of the different candidates supports a different mix of issues and policies, these dimensions define a latent "issues" space.

The position of each politician j in this p-dimensional space is defined by the factor loadings

λ_j; politicians positioned close to each other in the p-dimensional latent space appeal to the same precincts and therefore cannibalize voter shares from each other, similar to what happened in the 2002 French presidential elections where many left-wing candidates cannibalized Mr. Jospin's votes. With this factor structure, the proposed model avoids the proportional draw assumption (Kamakura & Russell, 1989), thereby producing predictions that take into account the differential effects of the eliminated candidates over the remaining ones.

Factor models such as the one proposed here are known to be invariant to rotation (Wedel & Kamakura, 2001). Therefore, the relative position of the candidates defined by the factor loadings (λ) and preference weights for the precincts can be directly interpreted as discussed above. On the other hand, due to this invariance any particular rotation is arbitrary, and therefore, interpretation of the latent dimensions themselves is highly subjective.

Finally, the coefficient of dissimilarity δ measures the extent to which candidates cannibalize each other's votes within each precinct, relative to abstentions. If the value of this coefficient is equal to 1, the elimination of a candidate in the second round will lead to an increase in the number of abstentions, as the original votes from the eliminated candidates will be distributed across all remaining alternatives, including abstentions. On the other hand, 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. Based on empirical evidence, we expect δ to be much smaller than 1.

3. Empirical illustration of the proposed voter-choice model

In order to demonstrate the features of the proposed model and test its predictive performance, we use the results from the 2002 elections for governor in the most important state of Brazil (São Paulo), for which we have complete voting counts for the two rounds at each of the 392 precincts in the state, covering all 25.6 million votes and abstentions. We also apply the model to the elections in 10 other states, but, due to space constraints, only report predictive performance results for them. We will use the data from the first stage to calibrate both the proposed model and a benchmark multinomial Logit model, and use the estimates from the first round to predict the final election outcomes.

3.1. Data description

The State of São Paulo is the most developed and populous of Brazil, representing about 22% of the Brazilian population. A total of 15 candidates participated in the 2002 elections for governor of São Paulo, representing coalitions between 27 political parties. However, four of these candidates had a negligible presence (less than 0.1% of the total vote) and were counted as abstentions in our analysis. The political campaign for governor started 4 months before the first-stage election on October 6, giving enough time for candidates to stake their positions and for voters to form their perceptions and preferences for the candidates. The two candidates receiving the most votes in the first stage participated in the second stage. Results from the first stage were known by the candidates 2 days after the polls closed, giving the remaining candidates 19 days to campaign until the final election on October 27, a relatively short period for the candidates to form alliances for the second round and to fine-tune their political platform.

The marketing effort by the candidates occurred in 3 forms: assemblies and contact with the voters in the streets; paid advertising on the radio and TV, outdoors, and in newspapers and magazines; and free campaign time on radio and TV. The schedule of free radio and TV advertising started 45 days before the first-stage election, corresponding to a total of 100 daily minutes on each TV network and 100 daily minutes on every radio station, distributed across all parties in proportion to party membership. For the second stage, free advertising occurred for only 11 days and was reduced to 80 daily minutes, equally distributed between the two candidates for president and two candidates for state governor. Finally, on the basis of the renderings of accounts registered by the candidates in the Electoral Regional Courts, more than 90% of the campaign expenses occurred in the first stage. This emphasis of the candidates on the first stage and the relatively short time between the two stages support the assumption implied by our Nested

Table 2
Data summary

	Share of votes			2nd round support	
	Total (%)	Average (%)	Std. dev. (%)	PSDB	PT
First stage					
Abstention	23.7	25.4	4.3		
PSDB	29.3	30.2	5.7	×	
PT	24.8	23.0	6.8		×
PPB	16.3	15.8	3.8	×	
PGT	2.7	2.3	1.2		×
PMDB	1.0	1.1	0.5		
PSB	0.8	0.8	0.3		×
PTB	0.8	0.8	0.5	×	
PV	0.2	0.2	0.1	×	
PRONA	0.2	0.1	0.1	×	
PSTU	0.1	0.1	0.1		×
PTC	0.1	0.1	0.0	×	
Second stage					
Abstention	20.2	21.5	3.6		
PSDB	46.8	45.9	5.8		
PT	33.0	32.6	6.7		

Demographics	Average	Std. dev.
Income	839.8	448.5
Famsize	3.5	0.2
% female	50.8	2.2
% 16 to 24 years	19.6	3.8
% 25 to 34 years	23.1	2.8
% 35 to 44 years	20.8	1.9
% 45 to 59 years	22.2	2.4

Logit Factor Model when predicting the final outcome based on first-stage votes, that voters' perceptions and preferences for the remaining candidates in the second stage are less likely to change substantially between the first and second stages than in the long campaign before the first stage.

Table 2 provides summary statistics for the available data. The first point to note in Table 2 is that the average share of votes across precincts is substantially different from the total share of votes, suggesting that there is significant heterogeneity in voter preferences across precincts. This is confirmed by the standard deviation of vote shares across precincts. One can also see from this table that the share of abstentions does not increase after candidates are eliminated in the first stage as one would predict with a multinomial Logit model. Moreover, this happens in virtually all precincts, suggesting that the differential draws between candidates and abstentions cannot be explained by voter heterogeneity, justifying our nested choice structure to separate the decision to vote and the choice of candidates in different nests. In fact, the share of abstentions decreases in the second stage, suggesting that voters are more motivated to go to the polls in the second stage when the governor is actually elected.

Our model estimates how candidates and voters' preferences are positioned in a latent space; the dimensionality of this space, however, has to be determined based on the data. Given that we have data for the whole population of voters in the state, rather than sampling data, the usual statistical criteria for selecting the number of dimensions are not appropriate. Therefore, we use cross-validation criteria for that purpose, based only on the first-round voting data; we fit our model to a randomly-selected sample of precincts, varying the number of latent dimensions from 1 to 5, and use the fitted models to make predictions for the remaining portion. Estimation of the proposed model between 1 and 5 latent factors led to the goodness-of-fit and cross-validation statistics reported in Table 3, indicating that the 3-factor solution is flexible enough to capture differences in preferences across precincts and parsimonious enough to provide the best expected predictions for the second round.

Parameter estimates for the proposed model are reported in Table 4. As one would expect from the total share of votes, there is considerable difference in general appeal (i.e., intercept) across the candidates, many of them being less attractive to the electorate than abstentions. The demographic coefficients also suggest that the candidates differ in their demographic appeal; candidates from PV, PSTU and PSDB are the most appealing and from PTB and PMDB are the least appealing to voters in precincts of high per-capita income. In a similar vein, precincts with a higher

Table 3
Selecting the number of factors based on holdout predictions

Number of factors	Log-likelihood	
	Calibration	Holdout
1	− 1,932,943	− 2,139,718
2	− 1,927,632	− 2,130,620
3	− 1,927,243	− 2,129,331
4	− 1,926,044	− 2,134,363
5	− 1,925,819	− 2,130,102

Table 4
Parameter estimates for the Nested Logit Factor Model Based on Votes from the First-round Election

Candidate	Intercept	Factor 1	Factor 2	Factor 3	Income	Famsize	% female	% 16–24 years	% 25–34 years	% 35–44 years	% 45–59 years
Abstentions	0.000	0.029	0.219	0.087	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PSDB	2.161	0.010	0.371	−0.105	0.094	0.132	0.105	0.340	0.020	−0.050	0.481
PT	1.879	−0.284	−0.235	0.038	0.008	−0.045	0.113	0.785	0.062	0.203	0.772
PPB	1.511	0.225	0.074	0.095	0.046	0.075	0.192	0.017	0.007	−0.090	0.252
PGT	−0.427	−0.048	−0.266	−0.034	0.008	0.091	0.199	0.638	0.180	0.147	0.614
PMDB	−1.166	−0.055	0.161	0.126	−0.068	−0.001	0.026	0.386	0.029	0.104	0.393
PSB	−1.459	−0.025	−0.043	−0.047	−0.054	0.090	0.136	0.481	0.122	0.066	0.508
PTB	−1.458	0.089	0.054	−0.263	−0.116	−0.033	0.083	0.244	0.122	−0.118	0.484
PV	−2.863	−0.025	−0.009	−0.006	0.107	0.118	0.082	0.431	0.022	0.065	0.561
PRONA	−3.337	0.140	−0.132	0.065	0.070	0.129	0.302	0.197	0.205	−0.087	0.424
PSTU	−3.409	−0.073	−0.129	0.041	0.104	0.162	0.224	0.754	0.186	0.037	0.893
PTC	−3.991	0.017	−0.064	0.003	0.068	0.109	0.203	0.355	0.074	−0.050	0.567
Coefficient of dissimilarity	0.361										

proportion of women see candidates from PRONA, PTSU and PTC more favorably and from PMDB, PB and PTB less favorably than the average voter. The fact that estimates for the *% female* predictor are positive for all candidates suggests that precincts with a higher proportion of female voters have lower abstentions than average.

Table 4 also shows an estimated dissimilarity coefficient of $\delta = 0.361$. As discussed earlier, this coefficient measures the extent to which abstentions compete directly with the candidates; a value close to 1 would suggest that abstentions compete directly with the candidates for votes. The fact that the estimate is substantially lower than 1 suggests that with the elimination of candidates in the first round, their voters are more likely to switch to one of the two remaining candidates rather than abstaining. Consequently, the main strategy for the two remaining candidates should be to increase their share of votes rather than trying to persuade citizens to cast their votes in the second round.

Figs. 1 and 2 show how our proposed model captures differences in voter perceptions and preferences across the 392 precincts. Fig. 1 displays the factor loadings for each candidate in the first round, so that candidates positioned close to each other tend to appeal to the same precincts and, therefore, compete more closely for their votes than with candidates away in the latent factor space. From Fig. 1, one concludes that the candidate from the PT party competed more closely with the PGT and PSTU parties in the first round of elections, (and thereby cannibalized his votes) while the candidate from the PSDB party competed more closely with the PPB and PRONA parties.

Fig. 2 shows the factor scores for each precinct as a directional vector. Candidates located farther from the origin in the direction pointed by a precinct's vector will have higher than average appeal to the voters in that precinct. For example, the precinct depicted by the solid vector in Fig. 2 has a higher-than-average attraction to the PT candidate, while the PSDB candidate has a higher-than-average appeal to the precinct depicted by the traced vector, after all other factors are taken into account.

The competitive positioning of the candidates (λ) in the first round (depicted in Fig. 1), along with the candidates' general (α) and demographic (β) appeal provide valuable insights regarding their strengths and weaknesses for the second round. Fig. 3 compares the demographic appeal (β) of the two remaining candidates for the second round, showing that the PT candidate is more attractive than the PSDB candidate among precincts with lower income, smaller families, and younger voters. These results seem reasonable, as the PT positions itself as the workers' party and the PSDB has a more conservative agenda catering to wealthier, older voters. Therefore, support from parties such as the PV and PSTU might help bring votes from these demographic segments to the PSDB. However, one must also consider whether this type of alliance is politically feasible. On one hand,

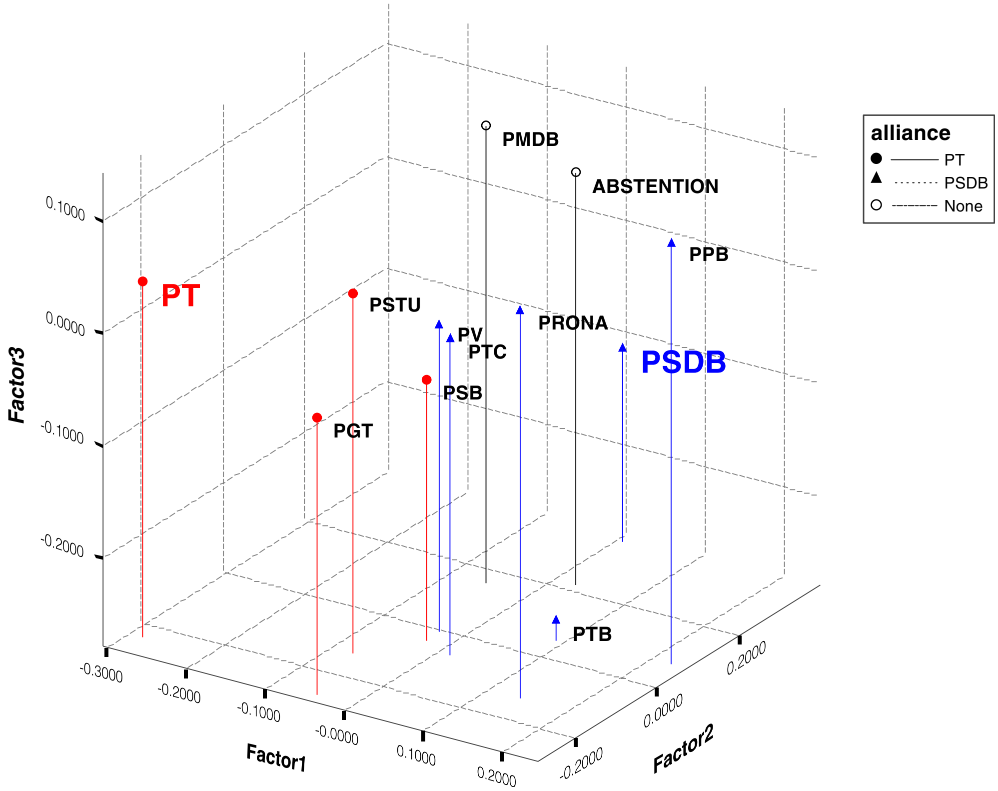

Fig. 1. Political positioning of the candidates.

alliances with candidates who cater to the same type of voter might not be necessary, as these voters will be naturally attracted to the remaining candidate with the elimination of their favorite one, given that they are likely to switch candidates, rather than abstain. On the other hand, alliances with politicians on the opposite side of the political spectrum may weaken the candidate's image among her core constituency.

The actual political alliances, announced in the media for the second round of elections, are depicted in Table 2 and Fig. 1. One can see that the two remaining candidates took a cautious approach, forming alliances with politicians who catered to similar voters in the first round and are therefore positioned closer to them in our positioning map than to the opposing candidate. However, one may wonder about the political "capital" gained through these alliances, given that our model suggests that voters from the eliminated candidates are more likely to switch to one of the remaining candidates rather than abstaining (i.e., $\delta \ll 1$). In this particular election, for example, the most critical ally among the eliminated candidates was the PPB candidate, who obtained 16.3% of the votes in the first round. However, it is unclear whether his support for the PSDB candidate in the second round would be of great value, because the PPB is the closest candidate to the PSDB (see Fig. 1), thereby suggesting that voters who chose the PPB will switch to PSDB in the second round, ceteris paribus.

The match between observed alliances and proximities in the latent spatial model suggests, however, that the positioning map generated by our model has good validity. In a more complex political environment, this map could also be used as a managerial tool to validate or identify less obvious political alliances.

3.2. Predicting the results of the second-stage election in São Paulo

If one assumes that voter preferences and the relative positioning of the remaining candidates are stable during the relatively short time between the first

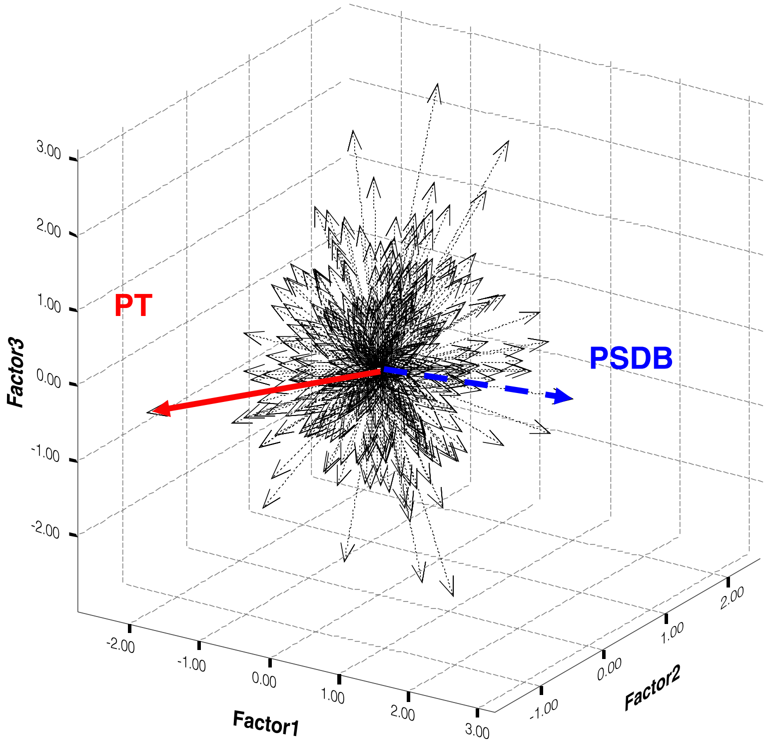

Fig. 2. Preference vectors for the precincts.

and second election stages, the model calibrated using the voting data from the first stage can be used to predict the final election results in each precinct. These predictions are conditional on the assumption that voter perceptions and preferences are the same as in the first round, and therefore should be used as "status-quo" predictions. However, these precinct-level predictions may be useful to the remaining candidates in planning how to best focus their political campaign in the few weeks left before the second round.

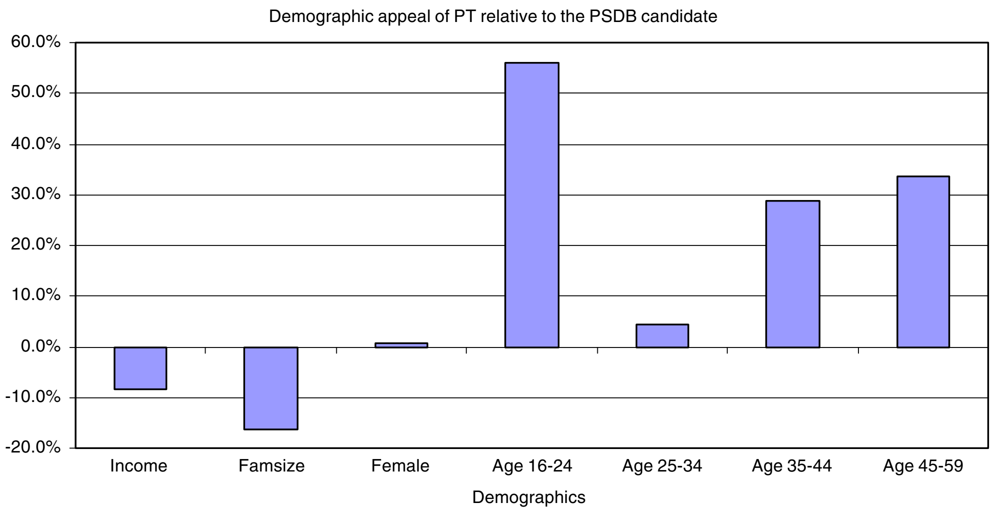

Fig. 3. Relative demographic appeal of the PT candidate.

Table 5
Predictive performance of the proposed and benchmark models in the second-stage elections (São Paulo)

Model	Mean absolute deviations			Average votes/precinct		
	ABST	PSDB	PDT	ABST	PSDB	PDT
Actual	–	–	–	13,204	30,635	21,609
Proposed Nested Factor Logit (with alliances)	3625	2216	3126	16,828	29,883	18,736
Proposed Nested Factor Logit (without alliances)	3820	4480	2464	15,368	27,152	22,927
Multinomial Logit	6720	6214	3554	19,921	24,631	20,895
Naïve	6745	6277	3571	19,933	24,634	20,880

The assumption of stability in voter preferences (our factor scores for each precinct Z_i and demographic coefficients β) and candidate positioning (λ) is critical in these predictions. However, because of the short time (median of 21 days across multiple countries) available for campaigning after the long campaign for the first round, perceptions and preferences are likely to hold, unless dramatic events or political scandals affect the remaining candidates and voter preferences. In the elections we studied, we also have strong evidence that most of the campaign resources were utilized before the first round of elections, again suggesting stable preferences and perceptions in the short time period between the two rounds. Nevertheless, the conditional forecasts produced by our model show how the two candidates will benefit within each precinct, from the elimination of the other candidates from the first round.

To assess and compare the predictive performance of our proposed modeling framework, we test two versions of our model against the multinomial Logit model, the most popular choice model in political science and other disciplines such as economics and marketing. We also use another benchmark, a "naïve" forecast assuming that the three alternatives remaining in the second round (winner, runner-up and abstentions) maintain the same share of votes observed in the first round, thereby ignoring the potential cannibalization of votes between these three alternatives and the eliminated candidates from the first round.[1]

The first version of our proposed model is the full version, as described earlier. The second version incorporates information about political alliances formed after the first round. This information is widely available in the media as politicians must broadcast their support for it to take any effect. In order to take the information about political alliances into account, we applied the following transformation to the candidates' intercepts

$$\alpha_k^* = \ln p_k + \frac{\ln \sum_{j=1}^{J} p_j - \ln p_0}{\delta}; \quad \alpha_0^* = 0 \tag{8}$$

where,

$$p_0 = \frac{e^{\alpha_0}}{e^{\alpha_0} + e^{\delta \ln \sum_{j=1}^{J} e^{\alpha_j}}}$$

$$p_k = \sum_{j=1}^{J} \left[\frac{e^{\alpha_j}}{\sum_{j'=1}^{J} e^{\alpha_{j'}}} (1-p_0) I_{jk} + \left(1 - \sum_{k'=1}^{J} I_{jk'}\right) \frac{e^{\alpha_k}}{\sum_{j'=1}^{J} e^{\alpha_{j'}}} \right]$$

$I_{jk}=1$ if politician j supports candidate k in the second election, 0 otherwise.

With the transformation above, the intercepts for the choice alternatives remaining in the second round take into consideration the full benefit from political alliances for the second election. Table 5 compares the mean absolute deviations for the vote forecasts produced with our proposed modeling framework and the two benchmarks.

Comparing these forecasts, which use only the estimates obtained from the first stage, one concludes that the two versions of our proposed modeling framework clearly outperform the naïve approach and the traditional multinomial Logit model. Our Nested Logit Factor Model avoids this proportional draw assumption made by the multinomial Logit and naïve models by considering a different degree of competition for candidates and abstentions, and accounting for unobserved heterogeneity in voter preferences, thus

[1] We thank an anonymous reviewer for suggesting this simple benchmark.

producing predictions that are closer to the actual votes than the other models. The fact that our factor model outperforms the multinomial Logit model shows that demographic alone are not sufficient to account for differences in political preferences across voters. By incorporating information about political alliances, the forecasting performance improves slightly for abstentions, and quite substantially for the PSDB candidate. However, the accuracy deteriorates for the PT votes.

For the two candidates involved in the final election, the most critical prediction is the margin of *valid* votes, relative to their opponent, which provides valuable information about her standing relative to the opponent in each electoral precinct. Based on this information, the candidate may decide how to focus her campaign in the few weeks left before the final election. For example, the candidate may decide to focus the campaign on precincts where the predicted share of valid votes is close to 50%, suggesting that voters might be more likely to change their minds before the second round. She might also take a defensive stand in precincts where her predicted share of votes is high. We compare the margin of votes for the PT candidate (relative to the PSDB) predicted by our final model, including the information about alliances, to the actual margin of votes (and share of

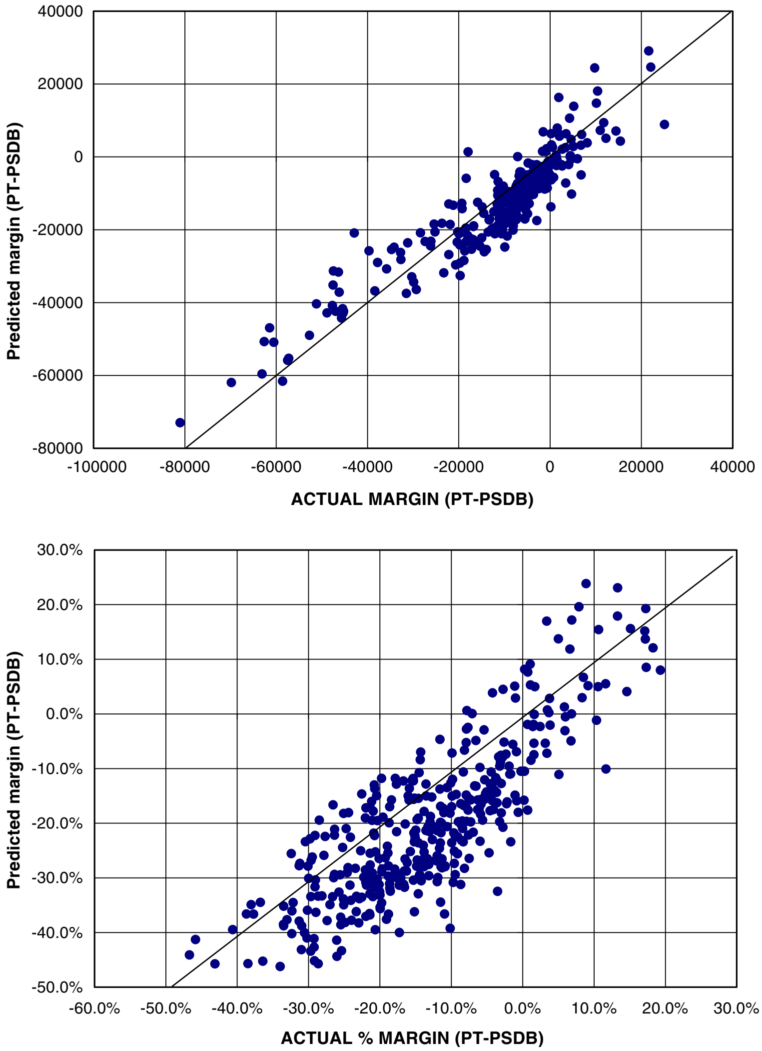

Fig. 4. Margin of votes and shares for the PT candidate.

votes) in Fig. 4. This performance suggests that the model is successful in identifying precincts where the PT candidate is likely to have a high margin, and where voting is likely to be evenly split between the two candidates.

4. Forecasting accuracy of the proposed model across multiple elections

In order to better verify the predictive performance of the proposed model, we applied it to the first-stage voting data for each of 11 state-governor elections in 2002 in Brazil where a second stage was required to choose between two runner-ups and enough precincts (more than 20) were available for model calibration, including São Paulo. The calibrated model from the first stage was then used to make voting predictions for the second stage at the precinct level. However, in this comparison of forecasting performance, we did not take political alliances into account because the information was not readily available to us for all of these elections. In Table 6, we compare the predictive performance of the proposed model with the same two benchmarks described earlier, in terms of the mean absolute deviation (MAD) computed across all precincts for the second-stage votes. These results show that, with only three exceptions (out of 33 measures), our proposed Nested-Logit Factor Model produces more accurate forecasts than the multinomial Logit and naïve models across the 1334 precincts in the 11 elections we studied.

While precinct-level forecasts are useful for strategic purposes, such as informing the candidates for political alliances and guiding the remaining candidates in their allocation of campaigning resources across precincts, the final election outcome is decided on the majority of votes; therefore, the ultimate forecast is the total number of votes. Table 6 compares the proposed model with the two benchmarks on this criterion. Once again, with a few exceptions (5 out of 33 performance measures), the proposed model produces forecasts of the total vote that are closer to actual than the two competing models. Table 7 also shows that the three models tend to over-predict abstentions (although the proposed model is clearly less biased than the two benchmarks). This happens because the actual percentage of abstentions decreased between the first and second rounds in most of the 112 elections we considered. Because the proposed and benchmark models are based on the voting shares observed in the first round, the elimination of any candidate can only lead to an increase in the shares for the remaining alternatives, thereby producing the over-prediction of abstentions. Note that because our proposed model allows for a different level of substitution among candidates and abstentions, it results in less bias than the benchmark models, even though it cannot predict a drop in the share of abstentions. Except for the

Table 6
Mean absolute deviation of vote forecasts across precincts for the second round of elections within eleven states

State	Precincts	Abstention				Winner				Runner-up			
		Actual votes/precinct	Mean absolute deviation			Actual votes/precinct	Mean absolute deviation			Actual votes/precinct	Mean absolute deviation		
			Proposed	Logit	Naïve		Proposed	Logit	Naïve		Proposed	Logit	Naïve
SP	392	13,204	**3820**	6723	6745	30,635	**4480**	6229	6277	21,609	**2464**	3581	3571
PR	206	8741	**752**	2243	2312	13,019	**2198**	3276	3422	10,587	2379	**1917**	2050
RS	173	7941	**1227**	2306	2550	18,201	**1276**	2033	2157	16,356	**1097**	1973	2343
CE	111	11,503	**2532**	4850	4828	15,907	3115	**1945**	2748	15,880	**5465**	6171	6317
SC	102	7978	**1110**	1907	2081	14,828	**2838**	3198	3293	14,625	**1959**	2399	2338
PA	87	12,334	**2898**	4026	4121	14,840	**2100**	2965	2785	13,853	**2156**	2862	3680
PB	76	7750	**1602**	2787	2853	11,710	**878**	2275	2364	11,094	**1336**	2730	2852
RN	68	8430	**1285**	2919	2957	12,067	**2381**	2789	2871	7700	1513	1331	**1224**
MS	52	6340	**610**	1080	1240	11,184	**572**	1110	1238	9626	**524**	926	1158
SE	35	7786	**3033**	5006	4976	13,756	**1207**	1743	1986	11,256	**2742**	3413	3675
RO	32	9164	**806**	3495	3447	10,878	**1864**	2417	2322	7538	**1629**	1671	2296

Values in bold show the lowest MAD.

Table 7
Aggregate forecasts for the final votes in eleven states

State	Abstentions				Winner				Runner-up			
	Actual votes/precinct	Predictions			Actual votes/precinct	Predictions			Actual votes/precinct	Predictions		
		Proposed	Logit	Naïve		Proposed	Logit	Naïve		Proposed	Logit	Naïve
SP	13,204	**15,368**	19,925	19,933	30,635	**27,152**	24,636	24,634	21,609	22,927	**20,887**	20,880
PR	8741	**8519**	10,935	10,951	13,019	**10,925**	9814	9728	10,587	12,903	**11,597**	11,668
RS	7941	**9065**	10,228	10,182	18,201	**17,511**	16,938	16,964	16,356	**15,921**	15,331	15,351
CE	11,503	**14,010**	16,354	16,306	15,907	18,864	17,225	**17,198**	15,880	**10,417**	9712	9787
SC	7978	**8738**	9850	9866	14,828	**12,269**	11,871	11,857	14,625	16,424	15,711	**15,708**
PA	12,334	**15,040**	16,339	16,461	14,840	**14,088**	13,327	13,352	13,853	**11,899**	11,361	11,214
PB	7750	**9228**	10,507	10,544	11,710	**11,484**	10,842	10,833	11,094	**9841**	9204	9177
RN	8430	**9664**	11,349	11,371	12,067	**10,111**	9281	9236	7700	8421	7567	**7589**
MS	6340	**6759**	7366	7373	11,184	**10,848**	10,528	10,533	9626	**9543**	9255	9243
SE	7786	**10,819**	12,792	12,761	13,756	**13,464**	12,160	12,131	11,256	**8515**	7847	7905
RO	9164	**9347**	12,659	12,611	10,878	**10,346**	8615	8893	7538	**7886**	6305	6075

Values in bold are the closest to the actual votes.

Parana (PR) and Santa Catarina (SC) states, all three models pick the winner in the final round. These two states are quite unique as they showed a dramatic reversal in votes between the two rounds. In Santa Catarina, the winner in the first round had 31.9% of the votes against 24.1% by the runner-up, but was narrowly defeated in the second round with 39.1% of the votes, compared to 39.6% by the other candidate. In Parana the reversal was even more dramatic (24.3% vs. 20.2% in the first round, and 32.7% vs. 40.2% in the second round).

5. Conclusions and directions for future research

The voter-choice model we proposed and tested in this study takes advantage of the preferences revealed by voters in the first round of a two-stage election, providing politicians with a mapping of their competitive position and of voter preferences in each precinct. Because the remaining politicians have limited time and resources with which to act between the first and second rounds, we assume that voter preferences and perceptions are reasonably stable during that short time period, and use the model calibrated on the first-round results to predict the final election outcome in each precinct.

Our tests of forecasting performance show reasonably good results, considering that we use only data from the first-stage elections to forecast the final outcome, without additional data such as historical indicators or opinion polls. Moreover, the performance is substantially better than two benchmark predictions (from the popular multinomial Logit model and a naïve approach) that ignore the cannibalization of votes among similar candidates in the first round. Most importantly, in addition to producing more accurate forecasts for the final elections at the precinct level, our model provides the political analyst and candidate with valuable insights about voter preferences and perceptions of the competing candidates to guide their campaign strategy. Our perceptual map (Fig. 1), inferred directly from observed voting behavior in the first round, shows how candidates were positioned relative to each other and to voter preferences in the first round. Our preference map (Fig. 2) indicates how each precinct differs, relative to the population at large, in its preference for the candidates. We believe forecasting is a critical and valuable input for decision making and planning; a forecasting model that provides insights about voter perceptions and preferences along with forecasting accuracy is even more valuable to decision makers.

Even though we account for the possible cannibalization of votes among similar candidates across precincts, our model uses the precinct as the unit of analysis, ignoring the diversity in voter preferences within each precinct. While the predictive performance we obtained was good, it can be improved by collecting voting data at a lower level of aggregation. For example, the precincts in São Paulo contain around 65,000 voters, on average, and therefore one would expect substantial diversity in preferences within each precinct.

In the eleven elections we studied, abstentions are generally lower in the second round relative to the first, when most of the candidates are eliminated, suggesting that the electorate is more motivated to cast their ballots in the final vote. This clearly violates the assumption of *regularity* (Tversky & Simonson, 1993) implied by random-utility choice models. In these choice models, choice probabilities will not decrease after alternatives are eliminated, contrary to what we find with abstentions. By nesting politicians in a different "branch" from abstentions in the nested Logit formulation, our proposed model, at the very best, predicts that abstentions will not increase substantially after some candidates are eliminated (depending on how small the dissimilarity coefficient δ is), but will never predict that abstentions decline in the second round. One potential solution to this problem would be the specification of a utility function that allows for loss aversion, so that each candidate is evaluated in relation to a reference point defined by the centroid of all politicians in the latent space. Such a utility function, for which losses (relative to the reference) loom larger than similar gains, may result in violations of regularity (Tversky & Kahneman, 1991), thereby predicting decreases in abstentions after the elimination of candidates. We attempted this, allowing the utility function in (1) to have different parameters for losses and gains, relative to the centroid of the latent space. However, because of the high level of aggregation within precincts, our results suggested loss proneness, rather than loss aversion, as an attempt by the model to capture within-precinct heterogeneity in preferences. Therefore, an extension of our model allowing for loss aversion would only be appropriate at a lower level of aggregation, when the assumption of homogeneity in preferences within each sampling unit would be more likely to hold.

To conclude, the use of demographic data to explain and predict voters' choices, and the identification of latent dimensions on which candidates compete against each other (and how they are positioned on these dimensions) offer interesting insights to guide political candidates' strategy for the remaining weeks before the second round of the election. Likewise, the analysis at the precinct level allows political candidates to identify the regions in which additional efforts or shifts in discourse are most likely to have the biggest impact, and hence could be used to allocate campaigning resources more efficiently between the two rounds of the elections, and optimally concentrate their efforts where they are more likely to swing the results to their advantage. Although this goes beyond the scope and focus of this paper, the forecasting model we propose should be seen as the first step toward the development of a decision-support system and optimization tool for political candidates, rather than a static forecasting instrument.

References

Armstrong, J. S. (2001). Combining forecasts. *Principles of forecasting* (pp. 417–439). Kluwer.

Bean, L. (1948). *How to predict elections*. New York: A.A. Knopf.

Berg, J., Forsythe, R., Nelson, F., & Rietz, T. (2000). Results from a Dozen Years of Election Futures Markets Research. *Working Paper*. Available at http://www.biz.uiowa.edu/iem/archive/BFNR_2000.pdf

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

Brown, L. B., & Chappell Jr., H. W. (1999). Forecasting presidential elections using history and polls. *International Journal of Forecasting*, *15*(2), 127–135.

Cho, S., & Endersby, J. W. (2003). Issues, the spatial theory of voting, and British general elections: A comparison of proximity and directional models. *Public Choice*, *114*, 275–293.

Cho, W. K. T. (1998). If the assumptions fit: A comment on the King Ecological Inference solution. *Political Analysis*, *7*, 143–163.

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

Coughlin, P. J., & Nitzan, S. (1981). Electoral outcomes with probabilistic voting and Nash social welfare maxima. *Journal of Public Economics*, *15*, 113–122.

Dasgupta, P., & Maskin, E. (2004). The fairest vote of all. *Scientific American*, *290*(3), 92–98.

Dow, J. K., & Endersby, J. W. (2004). Multinomial probit and multinomial Logit: A comparison of choice models for voting research. *Electoral Studies*, *23*, 107–122.

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., & Hinich, M. J. (1984). *The spatial theory of voting*. New York: Cambridge University Press.

Freedman, D. A., Klein, S. P., Ostland, M., & Roberts, M. (1998). On "solutions" to the Ecological Inference Problem. *Technical Report* No. 515, Statistics Department, UC Berkeley.

Gelman, A., & King, G. (1993). Why are American presidential election campaign polls so variable when votes are so predicable? *British Journal of Political Science*, *23*, 409–451.

Gourieroux, C., & Montfort, A. (1997). *Simulation based econometric methods*. Oxford University Press.

Greene, J. P. (1993). Forewarned before forecast: Presidential forecasting models and the 1992 elections. *PS, Political Science & Politics*, *26*, 17–21.

Holbrook, T. M., & DeSart, J. A. (1999). Using state polls to forecast presidential election outcomes in the American states. *International Journal of Forecasting*, *15*(2), 137–142.

Jérôme, B., Jérôme, V., & Lewis-Beck, M. S. (1999). Polls fail in France: Forecasts of the 1997 legislative elections. *International Journal of Forecasting*, *15*(2), 163–174.

Kamakura, W. A., & Russell, G. J. (1989, November). A probabilistic choice model for market segmentation and elasticity structuring. *Journal of Marketing Research*, *26*, 379–390.

King, G. (1997). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton: Princeton University Press.

Klein, J. G., & Ahluwalia, R. (2005). Negativity in the evaluation of political candidates. *Journal of Marketing*, *69*(1), 131–142.

Kramer, G. H. (1971). Short-term fluctuations in U.S. voting behavior, 1896–1964. *American Political Science Review*, *65*, 131–143.

Lewis-Beck, M. S., & Tien, C. (1999). Voters as forecasters: A micromodel of election prediction. *International Journal of Forecasting*, *15*(2), 175–184.

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

Paap, R., van Nierop, E., van Heerde, H. J., Wedel, M., Franses, P. H., & Alsem, K. J. (2005). Consideration sets, intentions and the inclusion of "don't know" in a two-stage model for voter choice. *International Journal of Forecasting*, *21*, 53–71.

Polly's Page (2005). http://morris.wharton.upenn.edu/forecast/Political/

Rallings, C., & Thrasher, M. (1999). Local votes, national forecasts—using local government by-elections in Britain to estimate party support. *International Journal of Forecasting*, *15*(2), 153–162.

Stambough, S. J., & Thorson, G. R. (1999). Toward stability of presidential forecasting: The development of a multiple indicator model. *International Journal of Forecasting*, *15*(2), 143–152.

Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge, UK: Cambridge Press.

Tufte, E. R. (1978). *Political control of the economy*. Princeton: Princeton University.

Tversky, A., & Kahneman, D. (1991, November). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, *106*, 1039–1061.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*, 1179–1189.

Wedel, M., & Kamakura, W. A. (2001). Factor analysis with mixed observed and latent variables in the exponential family. *Psychometrika*, *66*(4), 515–530.

Zeng, L. (2000). A heteroscedastic generalized extreme value discrete choice model. *Sociological Methods & Research*, *29*(1), 119–145.