Predicting the final outcome of two-stage elections with a voter-choice model

Wagner A. Kamakura
Duke University

José Afonso Mazzon
University of São Paulo

Arnaud De Bruyn
ESSEC Business School

Submitted: December 4, 2005

Proceedings of the 35th EMAC Conference
George J. Avlonitis, Nikolaos Papavassiliou and Paulina Papastathopoulou (Ed.)
Athens: European Marketing Academy

Predicting the final outcome of two-stage elections with a voter-choice model

The main purpose of this study is to propose a voter choice model that allows politicians to predict the final outcome of two-stage elections at the precinct level, based on the actual results from the first stage. 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 main research topic in the political science literature. Understanding voter preferences and forecasting the final outcome of elections is of critical value to politicians, as they can use the insights gained from the exercise to fine-tune their campaign strategies.

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. Recent debacles of opinion polls in predicting election outcomes in the US, in Britain (Rallings and Thrasher 1999), or in France (Jérôme, Jérôme and Lewis-Beck 1999, Dasgupta and Maskin 2004), have highlighted the limitations of such samplebased, stated-preference methods.

A review of the election forecasting literature shows that, "with few exceptions, forecasting studies have focused on the [US] presidential outcome" (Holbrook and 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, a typically large number of candidates compete in the first stage. Unless the candidate receiving the most votes reaches a minimum share of valid votes, a second stage election is held in which a subset of the candidates compete, within a few weeks from the first election. Two-stage elections also represent an invaluable opportunity to apply the theoretical and methodological advances in the modeling of voter choice. First, the results from the first stage, which are usually available within a few days from 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*=21days, *mean*=25 days, see Table 1), voters are unlikely to change their perceptions of the remaining candidates and even less likely to change their political preferences or leanings, 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 the insights about voter preferences to form strategic alliances for the second round. 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 finetune their campaign at the local level.

The voter-choice model we propose here attempts to overcome some of the challenges in predicting the outcome of two-stage elections, 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 voter decision to abstain in the election is different from the choice among the available candidates, thereby providing a better accounting of abstentions than the popular multinomial Logit formulation usually used in modeling voter choice.

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 follow the directional or vector perspective (Lishaug and Rabinowitz 2001) in the model we propose below. Following the probabilistic voting literature (Coughlin 1992; Coughlin and 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 (Zeng 2000; Dow and Endersby 2004; Paap, Nierop, Heerde, Wedel, Franses and Alsem 2005).

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,

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

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 as 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 and Russell 1989).

In developing this model, we also want to take into account that the decision to cast a vote is related, 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, we model the choice of candidate nested under the voting decision using a Multinomial Nested Logit formulation (Train 2002) with a dissimilarity coefficient δ, 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.

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. Table 2 provides summary statistics for the available data. The first aspect 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.

We estimated the proposed model with 1 through 5 latent factors, and based on crossvalidation statistics chose the 3-factor solution as 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 3. 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 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 3 also shows an estimated dissimilarity coefficient of δ=0.361. This coefficient measures the extent to which abstentions compete directly with the candidates; 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.

Figures 1 and 2 show how our proposed model captures differences in voter perceptions and preferences across the 392 precincts. Figure 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. Figure 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 Figure 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 Figure 1), along with the candidates' general (α) and demographic (β) appeal, provide valuable insights regarding their strengths and weaknesses for the second round. For example, we know that the PT candidate has a disadvantage relative to the other remaining candidate for the second round (PSDB) in precincts with high-income, and large-family voters, as shown in Table 2. Therefore, support from parties such as the PV and PSTU might help bring votes from these demographic segments.

The actual political alliances, announced in the media for the second round of elections, are depicted in Figure 1, where 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.

For the two candidates involved in the final election, the most critical prediction is the margin of *valid* votes, relative to their opponent. We compare the margin of votes for the PT candidate predicted by our final model (including the drop in abstentions and information about alliances) to the actual margin of votes in Figure 3, where one can see that the model tends to over-predict the margin of votes for the PT candidate.

Nevertheless, the proposed model is fairly effective at picking the winner across all precincts, correctly predicting the winner in 93.4% of the precincts (366 out of 392).

Conclusions

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 to act between the first and second rounds, we assume that voter preferences and perceptions are 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.

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. Therefore, we recommend that data be gathered at a lower level of aggregation (i.e., ballot box, or electoral zone) for implementation of our proposed model, for fine-tuned political decisions and more accurate election predictions.

References

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

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

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.*

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

Dow, J. K. and J. W. Endersby (2004) "Multinomial probit and multinomial logit: A comparison of choice models for voting research," *Electoral Studies*, 23, pp.107-22. Downs, A. (1957) *An economic theory of democracy*. New RYork: Harper & Row.

Holbrook, Thomas M. and Jay A. DeSart (1999) "Using state polls to forecast presidential election outcomes in the American states," *International Journal of Forecasting, 15(2), 137-142.*

Jérôme, Bruno, Véronique Jérôme and Michael S. Lewis-Beck (1999) "Polls fail in France : forecasts of the 1997 legislative elections," *International Journal of Forecasting, 15(2), 163-174.*

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structuring," *Journal of Marketing Research*, November, 379-90.

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

Lewis-Beck, Michael S. and Charles Tien (1999) "Voters as forecasters: a micromodel of election prediction," *International Journal of Forecasting, 15(2), 175-184.*

Paap, R., E. van Nierop, H.J. van Heerde, M. Wedel, P.H. Franses, K.J. Alsem (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.

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

Train, Kenneth E. (2003) *Discrete Choice Methods with Simulation*, Cambridge, UK: Cambridge Press.

Tufte, E. R. (1978) *Political Control of the Economy*, Princeton University, Princeton.

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.

Table 1 – Countries with elections in two turns by continent

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

(*) Typical number of days between the first and the second turn of the elections, when available.

Table 2 – Data Summary

	Share of votes			2nd round support	
First stage	Total	Average	Std Dev	PSDB	PT
Abstention	23.7%	25.4%	4.3%		
PSDB	29.3%	30.2%	5.7%	x	
PT	24.8%	23.0%	6.8%		x
PPB	16.3%	15.8%	3.8%	x	
PGT	2.7%	2.3%	1.2%		x
PMDB	1.0%	1.1%	0.5%		
PSB	0.8%	0.8%	0.3%		x
PTB	0.8%	0.8%	0.5%	x	
PV	0.2%	0.2%	0.1%	x	
PRONA	0.2%	0.1%	0.1%	x	
PSTU	0.1%	0.1%	0.1%		x
PTC	0.1%	0.1%	0.0%	x	
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		

Table 3 – Parameter Estimates for the Nested Logit Factor Model

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										

Figure 1 – Political positioning of the candidates

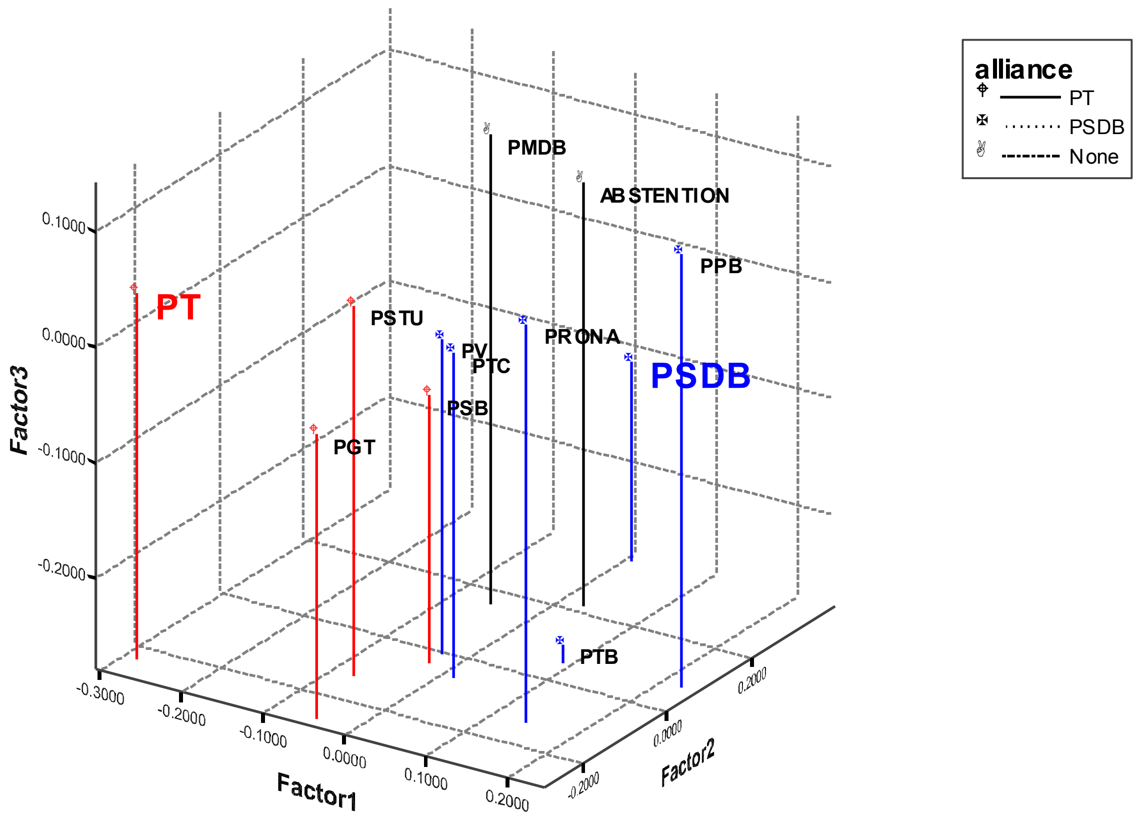

Figure 2 – Preference vectors for the precincts

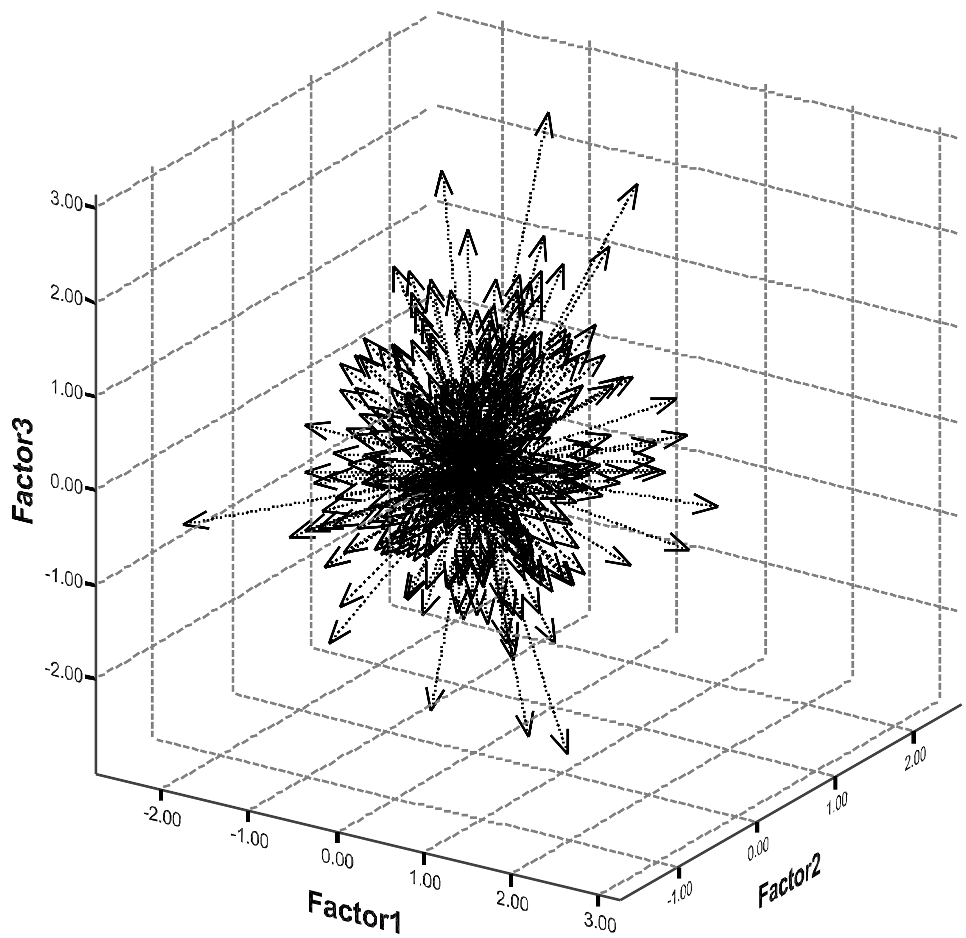

Figure 3 - Margin of votes and shares for the PT candidate