ESTIMATING REPRESENTATIVES FROM ELECTION POLL PROPORTIONS: THE SPANISH CASE

Jose M. Pavía[1]

Department of Applied Economics, University of Valencia, Valencia, Spain

Belén García-Cárceles

PhD Candidate, University of Valencia, Valencia, Spain

Elena Badal

PhD Candidate, University of Valencia, Valencia, Spain

Received: 17th February 2014/ Accepted: 30th November 2015

Abstract. *In a typical Spanish national or regional election poll, at most a couple of thousand electors are interviewed. This makes estimates of share of votes only reliable in aggregate terms. However, in multidistrict elections (as in the case of Spanish national elections) representatives are elected in subnational constituencies, where proportion estimates are untrustworthy. To bridge this gap, for the first time in the literature (as far as we know), this paper proposes some (initial exploratory) model-based approaches to solve the issue of translating poll proportions into seats in the context of the Spanish electoral system. The analyses performed in this study show that this avenue of research looks promising and that it should be further explored. The two approaches detailed in this paper substantially reduced the mean squared error (MSE) of direct forecasts, although still retaining an important part of the sampling variability.*

Keywords: *Surveys, Seat Allocation Forecast, Parliament Prediction, Model-Based Approach, Elections.*

1. INTRODUCTION

Opinion and election polls are ordinarily used in western democracies as a tool to assess policies and governments and as a medium to know and understand public desires. In almost all democracies voting intention between elections is regularly gauged through opinion polling, with surveys designed to forecast share of votes.

[1] Jose M. Pavía, email: Jose.M.Pavia@uv.es

However, what is relevant in any electoral contest is the number of representatives each political party earns.

Different proposals have been suggested in the literature to bridge the gap between proportion estimates and election outcomes, such as the so-called cube law that emerges in unimodal plurality systems (e.g., Gudgin and Taylor, 2012). Unfortunately, no universal solutions exist to map proportions of votes into representatives, as each electoral system has a specific electoral formula; surprisingly, no proposals can be found in the literature to solve this problem in the Spanish case. Indeed, the only works in the literature that deal in some way with this issue are Delicado and Udina (2001) and Udina and Delicado (2005)[2], who study (the origin of) the bias that the Spanish system introduce in polling forecasts, although they use extremely large samples sizes (~15,000). Our work starts to fill this gap by offering some initial exploratory answers to the issue of translating votes into seats in the context of the Spanish system.

In this paper non-sampling errors (including nonresponse bias, false reporting or missing data) that usually affect polls more seriously than sampling fluctuations (Pavía and García-Cárceles, 2012) are not addressed. We shall see that even under ideal sampling conditions the estimation of the composition of the Parliament from poll proportions is a challenging issue in which interesting and complex problems arise. Indeed, Díez-Nicolás (1996) considers that, independently of the quality of the survey data, the singular properties of the Spanish electoral system make it practically impossible to accurately predict the distribution of parliamentary seats among contending parties; Granados even claims that "[t]his deficiency is impossible to overcome" (Granados, 2005 p. 890). As we will see, this is far from true, and significant advances can be made by introducing in the forecasting process information from outside the sample.

The rest of the paper is organized as follows. Section 2 describes the (Congress) Spanish electoral system, which is also similar to the Parliamentary electoral systems of all the Spanish regions. In Section 3, we display (mainly using graphical methods) the extreme variability that direct forecasts from typical samples show and confirm the bias that emerges under the Spanish system in the Parliamentary composition forecasts. In Section 4, we propose a model to generate seat forecasts from proportion estimates which, despite substantially improving

---

[2] Despite the title of the paper "Predicciones de escaños electorales mediante encuestas" [Election seat forecasts from polls] (Díez Nicolás, 1996), in that work neither samples nor surveys are considered. On the other hand, Caballé et al. (2013) analyze the accuracy of published seat predictions against what they call a statistically "perfect poll".

direct sampling predictions, still exhibits too much variability. Section 5 goes deeper and suggests a superior model, which remains not completely satisfactory though. Finally, the last section concludes and points to future ways of improving the models introduced in this paper. An appendix about the d'Hondt rule closes the paper.

2. THE SPANISH (CONGRESS) ELECTION SYSTEM

Results of the 2012 Galician regional Parliamentary election and the 2011 Spanish general elections[3] are used to illustrate the performance of the two models detailed in this paper. In this section we describe the Spanish electoral system and some issues about the political circumstances concerning both electoral processes.

The Spanish Parliament has 350 members, elected for a maximum four-year term, determined by applying the d´Hondt rule to the votes gained by the party closed lists in each constituency (see Appendix). Spain is divided into 52 constituencies (50 provinces plus the cities of Ceuta and Melilla). Seats are allocated into constituencies as follows: (a) Ceuta and Melilla elect a seat each; (b) two seats are initially assigned to each of the 50 provinces; (c) the Hamilton rule (see, e.g., Pavía-Miralles, 2011, for details) is used to apportion the remaining 248 seats among the provinces, using provincial total populations as weights. The electoral law, which has remained stable since the reinstatement of democracy, enforces a compromise that over-represents the less populated provinces. Figure 1 displays on the map the number of seats per constituency for the 2011 elections. As can be observed, there is a large majority of medium or small districts; only Madrid and Barcelona have more than 30 seats.

Eleven national elections have taken place in Spain since democracy was re-established in 1977, and the party system has remained quite stable since then. The most significant change happened in the 1982 election, when the incumbent centrist party Union of the Democratic Centre (UCD) disappeared in practice, being replaced thereafter by the right-wing, conservative party People's Alliance (AP;

---

[3] Although in the Spanish general elections representatives of both the Congress of Deputies (350 seats) and the Senate (208 out of 266 seats) are elected, in this work we equate the Spanish general election with the Congress election. We omit the Senate elections in our research because they are (undoubtedly) of less interest.

[4] Of course during all those years some very small national and regional parties have been born and have died (sometimes after changing their acronyms), but the largest national and regional parties have remained the same over the years.

latter People's Party: PP).[4] From a political point of view, currently, there are three national parties – PP, the Socialist Party (PS) and the United Left (IU) – and a fourth (Union, Progress and Democracy: UPyD) emerging.[5] There are also several strong regional parties: Convergence and Union (CiU) and Republican Left of Catalonia (ERC) in Catalonia, Basque National Party (PNV) and Amaiur in the Basque Country, Galician National Bloc (BNG) in the Galician region and Canarian Coalition (CC) in the Canary Islands. Sometimes, other regional formations (PAR-CHA, PA, …) have also gained a seat. In the 2011 Spanish general elections thirteen parties (PP, PS, IU, UPyD, CiU, ERC, PNV, Amaiur, BNG, CC, Na-Bai, Q and FAC) achieved representation in the Congress.

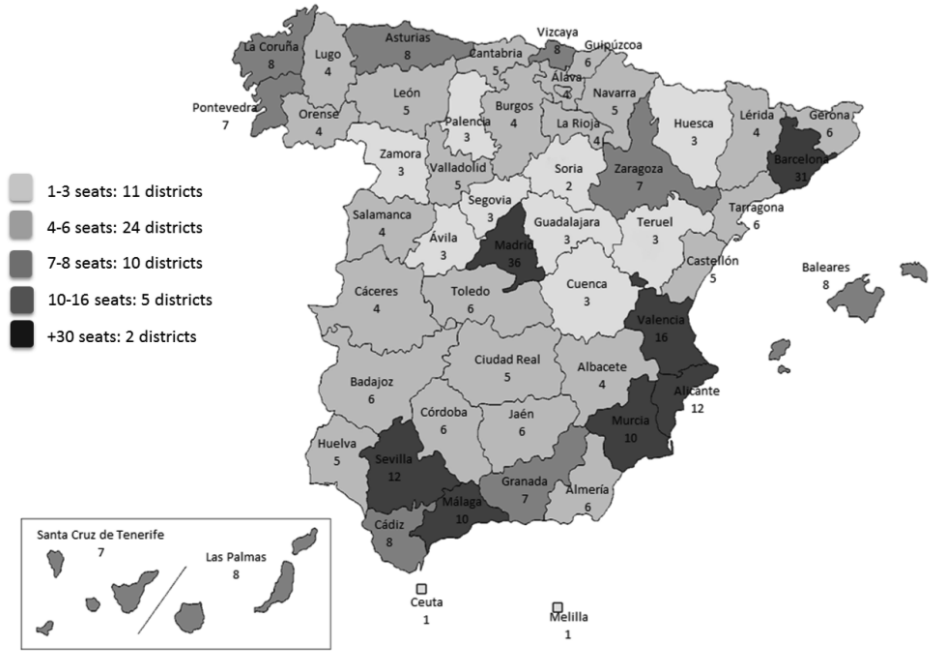

Figure 1: Map of the division of Spain in constituencies for the 2011 Spanish Congress general elections. The number of seats elected in each constituency and its corresponding Spanish name is provided on the figure.

[5] At the time of revising this paper, UPyD can disappear according to polls and two new national formations are emerging: Podemos was the fourth most voted party in Spain in the 2014 European Parliament elections, and Ciudadanos (Citizens) are in the way of becoming a major national party.

Galician 2012 Parliament elections are closely similar to Spanish Congress elections, with some obvious differences. It has only four constituencies (the Galician provinces: A Coruña, Lugo, Ourense and Pontevedra), to which are apportioned, respectively, 24, 15, 14 and 22 seats. The seats are again distributed independently in each constituency using the d'Hondt rule (see Appendix), but only among the parties that surpass the threshold of 5% of valid votes.[6] In the 2012 Galician Parliamentary elections four parties – PP, PS, BNG and Anova (a coalition of IU and a splinter group of BNG led by the former leader of BNG) – achieved representation.

3. EXTREME VARIABILITY OF DIRECT ESTIMATES IN DISTRICTS AND AGGREGATE BIAS OF SEAT ALLOCATION

In a typical Spanish national or regional election poll at most a couple of thousand electors are interviewed[7]. Election polls therefore just offer a general picture of the foreseen outcomes in the whole electoral space, with predictions of share of votes being reliable only in aggregate terms. However, in multidistrict elections (as in the case of Spanish national elections), representatives are elected in subnational constituencies; therefore a proper allocation of seats requires constituency-level predictions. Unfortunately, national proportion estimates are rarely representative of any constituency in particular. They are weighted estimates of the proportions in each constituency. And, moreover, poll estimates are not reliable at the constituency level.

For instance, if we consider the province of Toledo, an example of constituency not especially complex in statistical terms (it was *de facto* a two-party district, with 6 seats in the 2011 election), Figure 2 highlights the extreme variability that seat prediction exhibits with a national sample size of 2,500. A direct translation to seats of poll proportion estimates would yield an allocation error of one seat with probability 42.17% and a deviation of two seats with probability 2.13%.

---

6 The threshold in Spanish Parliamentary elections is 3%.

7 For example, considering the polls published in the last national electoral cycle (2008-2011) we found that on average 1,480 electors were interviewed per poll and that this average only grew to 4,726 during the last weeks of the campaign. The latter number is mainly a consequence of a single large survey conducted by the public institute CIS (Centro de Investigaciones Sociológicas) in its attempt to reach in all constituencies reliable proportion estimates on which to base a Parliamentary forecast.

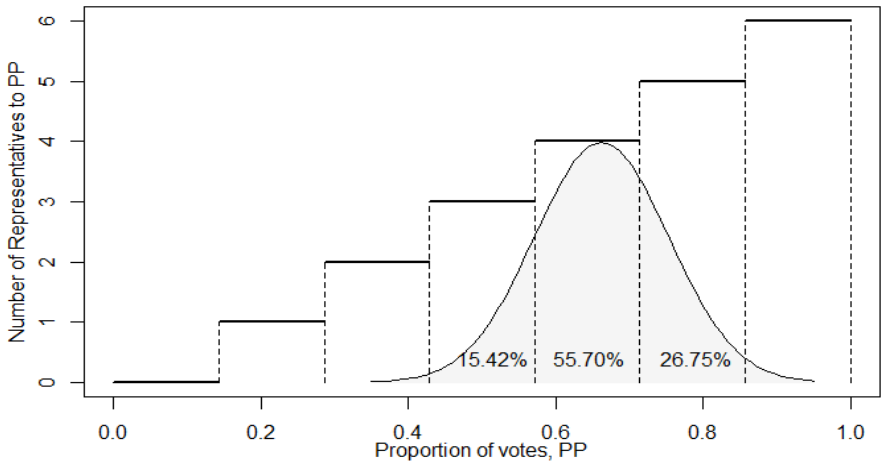

Figure 2: Example of seat allocations for the 2011 Spanish General Election in the province of Toledo (six seats) from constituency poll proportion estimates. In the 2011 General Election the two main parties' actual share of votes in Toledo was 66.22% for PP and 33.78% for PS. The horizontal axis is the proportion of votes for PP; PS has the rest. The vertical axis represents the number of seats that PP would gain under each proportion. The bell-shaped curve shows the approximate theoretical sampling distribution of the sample proportion for PP, assuming a national stratified sample size of 2,500 respondents with a proportional allocation of the sample size using constituencies as strata.

The use of province-level direct poll proportion estimates in all the constituencies would therefore produce significantly volatile Parliamentary predictions (see Figure 3), due to the small sample sizes collected in each constituency.

Furthermore, although when there are many constituencies we might expect the district biases to cancel each other out and the prediction for the whole Parliament to have no bias, data from real elections in Spain show that it is not usually the case (Delicado and Udina, 2001; Udina and Delicado, 2005), as for instance in the 2011 elections (see Figure 3). Indeed, direct poll empirical bias in the 2011 election for the number of seats allocated to the three main parties is, in aggregate terms, -1.5 for PP, 0.3 for PS and 1.7 for IU.

In our opinion, this aggregate seat bias that even ideal samples show in the Spanish case may occur due to (a) some locally important parties' competing only in a few districts and (b) similar biases consequence of national general voting patterns and the large number of small constituencies (see Figure 1).

The bias issue, however, does not occur with proportions, as can be observed in the biplot of Figure 4. As expected, poll proportions are unbiased, and furthermore they are less volatile, as the coefficients of variation (CV) clearly shows. The

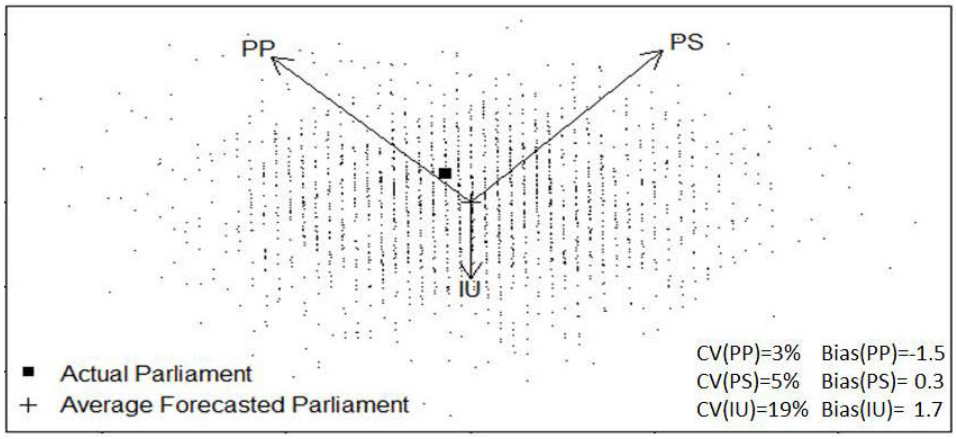

Figure 3: Biplot of Parliament Forecasts: High Variability and Bias. The points in the scatterplot represent 2000 Parliaments projected on the plane defined by the two main principal components obtained from 2000 simulated stratified polls of size 2,500 (with a fixed allocation of 15 units per district and the remaining 1,750 units allocated proportionally to district populations) using the 2011 Spanish General election final results. The two-component captured variance is 44.6%. The arrows, starting from the average point of forecasted Parliaments, represent the directions favouring the three main parties. The black square marks the position of the real Parliament and shows visually that there is a significant bias in the estimation of Parliamentary composition.

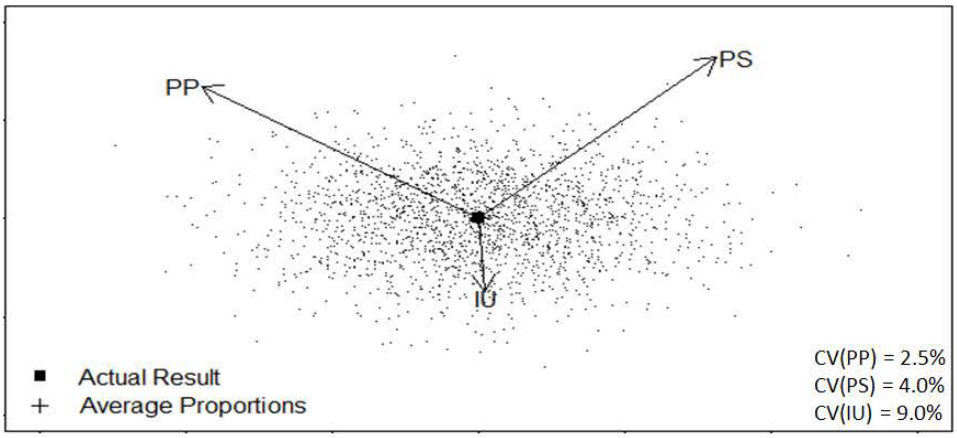

Figure 4: Biplot of Poll Proportions: Less Variability and No Bias. The points in the scatterplot represent 2000 Parliaments projected on the plane defined by the two main principal components obtained from 2000 simulated stratified polls of size 2,500 (with a fixed allocation of 15 units per district and the remaining 1,750 units allocated proportionally to district populations) using the 2011 Spanish General election final results. The two-component captured variance is 42.7%. The arrows, starting from the average point, represent the directions favouring the three main parties. The black square marks the position of the proportions and shows visually that there is no bias in the estimation of proportions.

Spanish electoral system, therefore, not only introduces bias in the process of translating votes into representatives, but it also amplifies forecast variability, mainly for small national parties, due to the average small constituencies. Indeed, as can be observed by comparing the CV for IU in Figures 3 and 4, the relative variability for IU in seat forecasts is more than twice that of proportions estimates.

4. MODEL 1: MODELLING THE VOTES-SEATS RELATIONSHIP

Although there is no universal votes-seats relationship in representative democracies, it could be argued that in a proportional system a party should receive around the same proportion of seats as its proportion of votes. It is, nevertheless, naïve to expect this. Indeed, as Gudgin and Taylor (2012, 2, Figure 1.1) illustrate analyzing 664 party results in 115 elections (Rae, 1971), proportional electoral laws are not always so directly concerned with that direct relationship. Despite this, they found a strong linear relationship between the proportions of votes and seats gained by the different parties. A similar linear pattern is obtained, in general terms, in the Spanish case. As can be inferred by Figure 5, a strong linear relationship exists between the proportions of votes and seats gained by parties in Spanish elections.

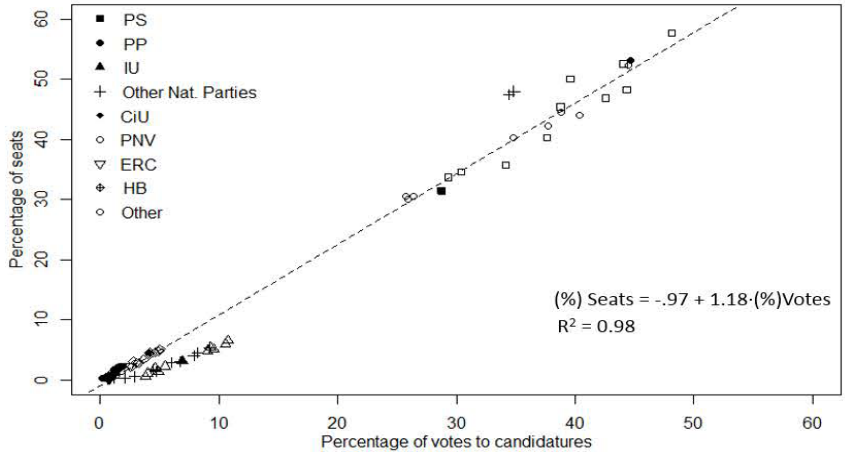

Figure 5: Historical relationships between percentage of votes and percentage of seats gained for the main Spanish parties in Congress elections (1977-2011). Solid symbols correspond to 2011 election outcomes. Eleven elections considered and 128 points drawn.

Despite the strong overall relationship ($R^2 = 0.98$), it seems that a single equation for each party could yield better results. For small national parties, the linear relationship seems to work only if they earn at least 4% of votes (see Figure 6); perhaps a logit transformation may be useful for the entire range (see Figure 6-left). On the other hand, for small regional and *sporadic* parties, as well as for the possible existence of selection bias (only parties reaching Parliament are displayed in the figure), an ordinal logit model with additional covariates could be more adequate (Figure 6-right).

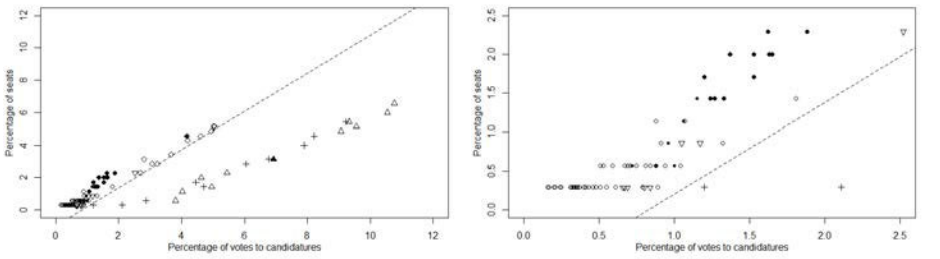

Figure 6: Details of parts of Figure 5 amplified. In the left panel, detail for small national parties. In the right-panel, detail for small regional and *sporadic* parties.

Thus, in light of the patterns we can infer from Figures 5 and 6, we propose a simple first model (Model 1) defined by the following rules: (a) fit independent univariate linear equations for PP and PS; (b) fit a univariate linear equation for small national parties (range 4-12% of national vote); (c) fit a univariate linear equation for Catalonian regional parties; (d) fit a univariate linear equation for regional parties from the Basque Country; (e) fit a univariate linear equation for other regional parties; (f) round to the nearest integer (except for point five predictions that are rounded to zero) the seat forecasts predicted with the linear fits; and (g) assign the difference to 350 seats to the majority party, given that the Spanish system favours the party obtaining the greatest support.

In order to assess the practical value of Model 1, we have (a) simulated[8] (assuming absence of non-sampling error and under simple random sampling in each constituency) 2,000 polls of size 2,500 (with a fixed allocation of 15 units per district and the remaining 1,750 units allocated proportionally to district populations)

[8] All the computations and graphical representations made in this paper have been performed in version 3.0.0 of the free statistical software R (R Core Team, 2013).

from the 2011 Spanish General election final results and (b) compared the Parliaments that would be obtained applying the d'Hondt rule in each constituency to the direct poll proportion estimates with the Parliaments that would be obtained applying Model 1 to the national poll proportion estimates.[9] A summary of the results attained is presented in Table 1: the use of the votes-seats relationship improves overall forecasts, and indeed it reduces the total mean squared error (MSE) by 22.4 percentage points.

Table 1: Comparison of Poll and Model 1 Parliamentary Forecasts (350 seats).

	Actual	**Direct Poll Forecasts**				**Model 1 Forecasts**			
Party	**Results**	Average	Bias	Std. Dev.	MSE	Average	Bias	Std. Dev.	MSE
PP	186	184.5	-1.5	5.5	32.4	187.5	+1.5	4.6	23.6
PS	110	110.3	+0.3	5.4	28.8	111.1	+1.1	4.9	24.9
IU	11	12.7	+1.7	2.5	8.9	11.9	+0.9	1.7	3.8
UPyD	5	6.5	+1.5	1.8	5.6	6.0	+1.0	1.5	3.2
CiU	16	16.7	+0.7	2.1	4.9	14.4	-1.6	1.9	6.2
ERC	3	2.6	-0.4	1.1	1.4	2.4	-0.6	1.1	1.5
PNV	5	5.2	+0.2	1.3	1.6	5.4	+0.4	1.3	1.8
Amaiur	7	5.3	-1.7	1.4	4.9	5.7	-1.3	1.3	3.5
BNG	2	1.5	-0.5	1.0	1.1	1.6	-0.4	0.6	0.5
Na-Bai	1	0.5	-0.6	0.5	0.6	0.1	-0.9	0.3	0.9
CC	2	2.0	0.0	1.0	1.0	1.2	-0.8	0.5	0.9
Q	1	0.7	-0.3	0.7	0.5	1.0	0.0	0.6	0.3
FAC	1	1.1	0.1	0.7	0.5	0.8	-0.2	0.5	0.3
PA	0	0.1	0.1	0.2	0.1	0.5	0.5	0.5	0.6
PRC	0	0.4	0.4	0.5	0.4	0.1	0.1	0.3	0.1

5. MODEL 2: MODELLING NATIONAL-PROVINCE VOTE RELATIONSHIPS

Despite the improvement that entails using the votes-seats model, Model 1 is quite sensitive: it translates a significant portion of the variability of poll proportion estimates to seat estimates. Hence, we follow a new approach in this section. In

[9] Six linear regressions have been estimated to implement Model 1. Eight data points were used for the PP equation, ten data points were used for the PS equation, fifteen for small national parties, twenty for Catalonian parties, twenty-six for Basque parties and thirty for the remaining regional parties.

particular, although there is still room to improve Model 1 – for instance, by modifying its rules and/or by reducing poll proportion variabilities using ratio estimators, post-stratification techniques or superpopulation models (Mitofsky and Murray, 2002; Mitofsky, 2003, Pavía and Larraz, 2012) –, we explore the impact of using global constituencies-votes relationships, see Figure 7.

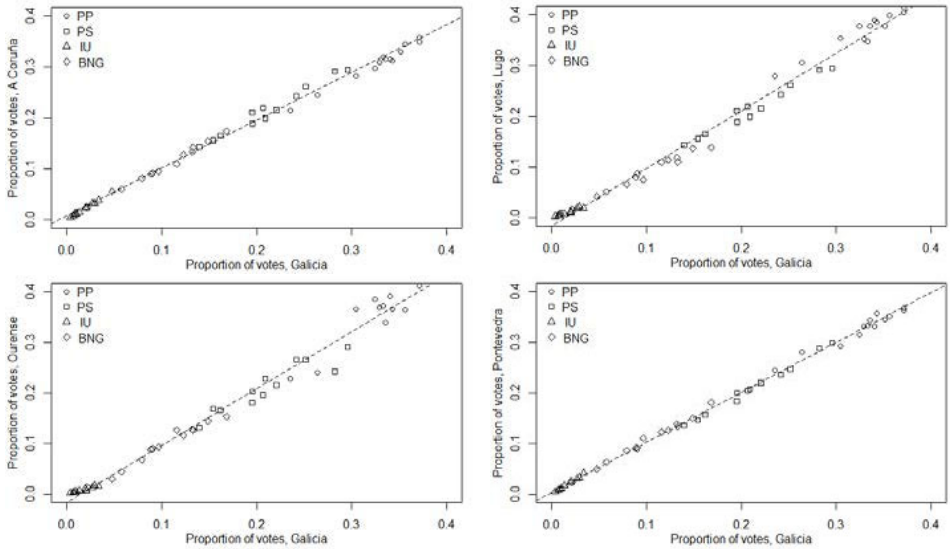

Figure 7: Global constituency-vote relationships for the constituencies of Galicia (1989-2009) in regional and general elections. Thirteen elections considered and 50 points drawn in each panel.

As can be observed in Figure 7, where a sample of the regional-province relationships of percentages of votes for Galician elections have been displayed, strong linear patterns exist between the proportions of votes a party reaches in the whole electoral space and in each constituency. Thus, we propose Model 2 that is defined by the following list of rules: (a) fit univariate linear models per each party and constituency between the total proportion of votes and the corresponding party-constituency proportion of votes; (b) use the proportion estimates obtained for each party in each constituency in (a) to allocate seats in the corresponding constituency; and (c) use for *new* parties the relationship of the most ideologically close existing party.

Table 2: Comparison of Poll and Model 2 Galician Parliamentary Forecasts.

	Actual	Direct Poll Forecasts				Model 2 Forecasts			
Party	**Results**	Average	Bias	Std. Dev.	MSE	Average	Bias	Std. Dev.	MSE
PP	41	40.8	-0.2	2.2	4.8	41.0	0.0	2.2	4.8
PS	18	17.5	-0.5	1.9	3.9	17.7	-0.3	1.6	2.6
Anova	9	9.8	+0.8	1.6	3.3	9.3	+0.3	1.4	1.9
BNG	7	6.9	-0.1	1.4	2.1	7.0	0.0	1.4	1.8

In order to assess the empirical value of Model 2, we have applied it to the 2012 Galician regional elections and the 2011 Spanish general elections[10], the latter with two different sampling designs. In Table 2 we report the results for the Galician case. In particular, to gauge the model in the 2012 Galician elections we have (a) simulated (assuming absence of non-sampling error and under simple random sampling in each constituency) 2,000 polls of size 800 (with a fixed allocation of 100 units per district and the remaining 400 units allocated proportionally to district populations) the 2012 Galician regional Parliamentary election final results and (b) compared the Parliaments that would be obtained applying the d'Hondt rule in each constituency to the direct poll proportion estimates with the Parliaments that would be obtained applying Model 2 to the aggregate regional poll proportion estimates. In this case, despite the larger number of seats allocated per constituency, which makes model improvements less difficult and evident, an advance in both bias and variance is obtained, with a reduction of MSE of 22.1 points.

In Table 3 the same polls used in Table 1 (as well as in Figures 2 and 3) were analyzed using Model 2 (this makes comparisons between models easier); an additional improvement is obtained with this new approach. In particular, the global reduction of MSE grows in this case to reach 29.0 points. It seems that the additional computational and logistic burden entailed by passing from estimating 15 regressions (Model I) to fitting 780 models (Model II) yields its benefits.

Further scrutiny of Table 3 outcomes, nevertheless, shows that in relative terms regional parties account for an important part of the total MSE. So, to end these exploratory analyses, we investigate the performance of Model 2 with an alternative sampling design in which an extra polling effort is made in those

[10] In the case of Galicia, the regressions have been estimated using the outcomes recorded in national or regional elections held from 1989 to 2011. Thirteen points have been used in PP and BNG equations and twelve in PS and Anova elections because PS and IU concurred in coalition in the 1997 regional elections. In the case of Spain, the models were fitted using the results recorded in national and European elections held since 1989 to 2009. Twelve points were used in every single regression.

constituencies where there is a historical strong presence of regional parties. In particular, to obtain the results reported in Table 4 we have simulated (assuming absence of non-sampling error and under simple random sampling in each constituency) 2,000 polls of size 2,500 (with a fixed allocation of 5 units per district in those constituencies without a history of regional parties and 30 units in districts with history of regional parties, with the remaining 1,890 units allocated proportionally to district populations) using the 2011 Spanish General election final results. In this case, a larger reduction of MSE is observed. The reduction now reached 37.6 points. Unfortunately, a comparison of Tables 3 and 4 suggests this is more a consequence of a worsening of direct poll forecasts than an actual improvement.

Table 3: Comparison of Poll and Model 2 Parliamentary Forecasts.

	Actual	**Direct Poll Forecasts**				**Model 2 Forecasts**			
Party	**Results**	Average	Bias	Std. Dev.	MSE	Average	Bias	Std. Dev.	MSE
PP	186	184.5	-1.5	5.5	32.4	188.4	2.4	4.4	24.9
PS	110	110.3	0.3	5.4	28.8	110.0	0.0	4.7	21.9
IU	11	12.7	1.7	2.5	8.9	11.0	0.0	1.8	3.4
UPyD	5	6.5	1.5	1.8	5.6	3.5	-1.5	1.0	3.2
CiU	16	16.7	0.7	2.1	4.9	16.4	0.4	1.4	2.0
ERC	3	2.6	-0.4	1.1	1.4	2.5	-0.5	0.8	1.9
PNV	5	5.2	0.2	1.3	1.6	5.7	0.7	1.3	1.8
Amaiur	7	5.3	-1.7	1.4	4.9	5.9	-1.1	1.2	3.2
BNG	2	1.5	-0.5	1.0	1.1	1.6	-0.4	1.4	0.7
Na-Bai	1	0.5	-0.6	0.5	0.6	0.5	-0.5	0.5	0.5
CC	2	2.0	0.0	1.0	1.0	2.7	+0.7	0.9	1.2
Q	1	0.7	-0.3	0.7	0.5	0.6	-0.4	0.6	0.5
FAC	1	1.1	+0.1	0.7	0.5	0.9	-0.1	0.6	0.3
PA	0	0.1	+0.1	0.2	0.1	0.0	0.0	0.0	0.0
PRC	0	0.4	+0.4	0.5	0.4	0.4	+0.4	0.5	0.4

The existence of strong subnational parties in Spain subtracts accuracy from the forecasts obtained with the proposed models, adding extra complexities to the issue which cannot be necessarily addressed in extra polling of constituencies with a strong presence of regional parties. One of the problems related to the presence of regional parties is the fact that differential turnouts among constituencies can have a significant impact on the performance of the models, hence some modifications of the proposed models in order to account for this could lead to a global improvement.

Table 4: Comparison of Poll and Model 2 Parliamentary Forecasts. Alternative sampling design.

	Actual	Direct Poll Forecasts				Model 2 Forecasts			
Party	**Results**	Average	Bias	Variance	MSE	Average	Bias	Variance	MSE
PP	186	183.2	-2.8	5.8	41.1	188.2	+2.2	4.5	25.3
PS	110	111.1	+1.1	5.6	32.3	110.2	+0.2	4.7	22.0
IU	11	12.9	+1.9	2.5	10.0	10.9	-0.1	1.9	3.5
UPyD	5	6.8	+1.8	1.8	6.5	3.5	-1.5	1.0	3.3
CiU	16	16.8	+0.8	1.9	4.1	16.4	+0.4	1.2	1.6
ERC	3	5.2	+0.2	1.1	1.3	5.7	+0.7	1.2	1.8
PNV	5	5.4	-1.6	1.2	1.4	5.9	-1.1	1.0	1.5
Amaiur	7	1.5	-0.5	1.3	4.3	1.7	-0.3	1.3	3.0
BNG	2	2.6	-0.4	0.8	1.0	2.5	-0.5	0.6	0.5
Na-Bai	1	0.4	-0.6	0.5	0.6	0.5	-0.5	0.5	0.5
CC	2	2.0	0.0	0.9	0.7	2.7	+0.7	0.8	1.1
Q	1	0.7	-0.3	0.6	0.5	0.6	-0.4	0.5	0.4
FAC	1	1.0	0.0	0.7	0.5	0.9	-0.1	0.6	0.4
PA	0	0.1	+0.1	0.2	0.1	0.0	0.0	0.0	0.0
PRC	0	0.4	0.4	0.5	0.4	0.3	0.3	0.5	0.3

Source: Own Elaboration.

6. CONCLUSIONS AND FURTHER REMARKS

Within the Spanish electoral context, two model-based approaches to translate poll proportions into seats have been proposed in this research. The suggested approaches significantly reduce the sampling variability associated with polls (22.1 points in Model 1, and 29.0 and 37.6 points in Model 2), although their forecasts still show much volatility (and this without accounting for model uncertainty).

Despite the easier relationships (see Figure 5) that link votes and seats in proportional systems, the high sensitivity that seat share exhibits along with small variations in share of votes makes this problem a difficult and interesting challenge.

The analyses performed in this paper show that this avenue of research looks promising and that there is still room for improvement. In particular, within the proposals presented here, both models deal with linear univariate relationships, hence, for example, they could be extended to multivariate versions (using, for example, SURE models) in which historical correlations can be taken into account. On the other hand, among other approaches, possible extensions that also deserve attention would include (a) considering the spatial dimension of the data, (b) using small area estimation models in which shrinkage constituency estimates are obtained combining model-based predictions and constituency poll estimates, or (c) modelling the relationships through multinomial models.

APPENDIX: THE D'HONDT RULE

The d'Hondt law is a particular case of a divisor rule that attempts to make the averages between votes received and seats obtained similar between parties.

Given K parties obtaining $p_1, p_2, \ldots, p_K$ proportion of votes and M seats to allocate, the d'Hondt rule proceeds as follows: (a) calculate the $K \times M$ matrix of quotients p_k/d_j, $k = 1, \ldots, K, j = 1, ..., M$, with $d_j = j$; (b) select the M largest quotients and give the corresponding parties a seat for each of their largest quotients.

Note that depending on the sequence of denominators d_j chosen, different rules emerge, such as the first-past-the-poll and the winner-take-all rules ($d_j = 1$) or the Sainte-Lagüe ($d_j = 2j$-1) and the modified Sainte-Lagüe rules.

Thus, if, for example, 7 seats must be distributed among four parties (A, B, C, and D) receiving 45%, 32%, 15%, and 8% of the valid vote, we must construct the matrix below, from which A, B, C, and D would obtain 4, 2, 1, and 0 seats (with the seat allocating order in super-indexes).

	Seats						
Party	1	2	3	4	5	6	7
A	**45.0**^{1}	**22.5**^{3}	**15.0**^{5}	**11.3**^{7}	9.00	7.50	6.40
B	**32.0**^{2}	**16.0**^{4}	10.70	8.00	6.40	5.30	4.60
C	**14.0**^{6}	7.50	4.70	3.50	2.80	2.30	2.00
D	8.00	4.00	2.70	2.00	1.60	1.30	1.10

As can be observed, the d'Hondt rule provides a proportional system that enhances the representation of the larger parties to the detriment of smaller ones. For instance, in the example an exact proportional distribution would lead to 3.15, 2.24, 1.05 and 0.56 seats for each party. Indeed, if only one seat is allocated, the d'Hondt algorithm reduces to a majority rule. Obviously, as the interested reader could check, the correction becomes smaller as the number of seats increases, so almost perfect proportional representation can be achieved if the number of seats is sufficiently large.

ACKNOWLEDGEMENTS

Some preliminary versions of this paper were presented at the 2013 EPOP Conference (Lancaster, United Kingdom) and the 2013 SESM Conference (Barcelona, Spain). The authors acknowledge the valuable comments of the meetings' discussants. The authors also thank Marie Hodkinson for revising the English of the paper and two anonymous referees for their really valuable comments

and recommendations. The support of the Spanish Ministry of Economics and Competitiveness through project CSO2013-43054-R is acknowledged. Belén García-Cárceles also acknowledges the "Atracció al Talent" scholarship provided by the Research Vicerrectorado of the University of Valencia. The usual disclaimers apply.

REFERENCES

Caballé, A., Grima, P. and Marco-Almagro, L. (2013). Are election polls right? Analysis of the accuracy of election poll predictions published in the press. *Revista Española de Investigaciones Sociológicas* 142: 25-46.

Delicado, P. and Udina, F. (2001). ¿Cómo y cuánto fallan los sondeos electorales?, *Revista Española de Investigaciones Sociológicas* 96: 123-150.

Díez Nicolás, J. (1996). Predicciones de escaños electorales mediante encuestas, *Revista Española de Investigaciones Sociológicas* 74: 269-289.

Granados, F.J. (2005). Intertwined cultural and relational environments of organizations, *Social Forces* 83: 883-918.

Gudgin, G. and Taylor, P.J. (2012). *Seats, Votes, and the Spatial Organisation of Elections*, ECPR Press, Colchester.

Mitofsky, W.J. and Murray, E. (2002). Election night estimation, *Journal of Official Statistics* 18: 165-179.

Mitofsky, W.J. (2003). Voter news service after the fall, *Public Opinion Quarterly* 67: 45-58.

Pavía, J.M. and Larraz, B. (2012). Nonresponse bias and superpopulation models in electoral polls, *Revista Española de Investigaciones Sociológicas,* 137: 237-264.

Pavía, J.M. and García-Cárceles, B. (2012). Una aproximación empírica al error de diseño muestral en las encuestas electorales del CIS, *Metodología de Encuestas* 14: 45-62.

Pavía-Miralles, J.M. (2011). Tamaño del congreso y resultados electorales, *Cuadernos "Manuel Giménez Abab"* 2: 84-96.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: Vienna (Austria). http://www.R-project.org/

Rae, D.W. (1971). *The Political Consequences of Electoral Laws*, Yale University Press: New Haven.

Rae, D.W. and Ramírez, V. (1993). *El Sistema Electoral Español: Quince Años de Experiencia*, McGraw-Hill: Aravaca.

Santoyola, P. (2011). III Jornadas de derecho parlamentario. La reforma de la ley orgánica del régimen electoral general, *Cuadernos "Manuel Giménez Abab"* 1: 95-98.

Udina, F. and Delicado, P. (2005). Estimating parliamentary composition through electoral polls, *Journal of the Royal Statistical Society A* 168: 387-399.

Urdanoz, J. (2008). El maquiavélico sistema electoral español, *El País*, 16 February, 35.