EPOP 2013
Elections, Public Opinion and Parties

LANCASTER UNIVERSITY

A statistical model to transform election poll proportions into representatives: The Spanish case

Jose M. Pavia

Elections and Public Opinion Research Group

Universitat de Valencia

13-15 September 2013, Lancaster University

Introduction

Opinion and election polls are ordinarily used in western democracies as a tool to know and understand public desires and to assess policies and governments.

Polls are usually designed to forecast share of votes, although election results are mediated for the particular electoral formula each system uses.

Introduction

Different proposals have been suggested to bridge the gap between estimates proportions and election outcomes, such as the so-called cube law that emerges in unimodal plurality systems (e.g., Gudgin and Taylor, 2012).

This work tries to offer some answers to the issue of translating votes into seats in the context of the Spanish elections.

Gudgin and Taylor, 2012, *Seats, votes, and the spatial organisation of elections*, ECPR Press.

The Spanish (Congress) Election System

- The Spanish Parliament has 350 seats.
- Spain is divided into 52 constituencies
  (50 provinces plus the cities of Ceuta and Melilla).
- Ceuta as well as Melilla elect a seat each.
- At least 2 seats are elected in each province.
- The Hamilton rule is used to apportion the remaining 248 seats among the provinces, using province total populations as weights.
- Votes are casted to party closed lists.

The Spanish (Congress) Election System

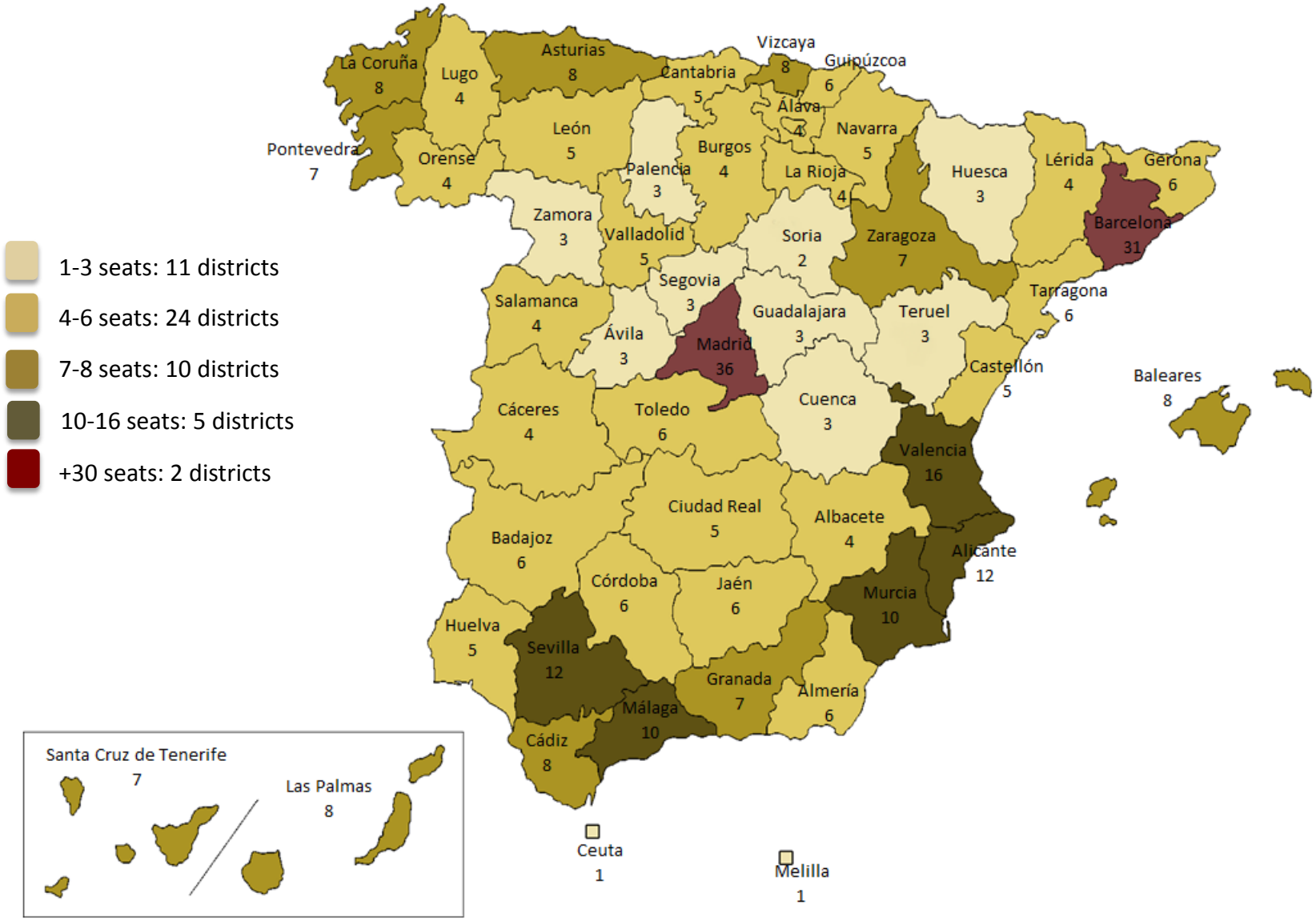
- Within each constituency, seats are allocated to parties using the d'Hondt rule.
- To opt to representation in a constituency, a party needs to reach at least 3% of the valid votes of the constituency.
- There are three national parties (PP, PS and IU) and a fourth (UPyD) emerging.
- There are several strong regional parties. CiU and ERC in Catalonia and PNV and HB-Amaiur in The Basque Country. BNG and CC and sometimes PAR-CHA, PA,… also used to reach representation.

The Spanish (Congress) Election System

According to many commentators (e.g., Rae and Ramírez, 1993; Urdanoz, 2008; Santayola Machetti, 2011), the Spanish system, being nominally a proportional system, *de facto* favours big parties as a consequence of the cumulative effect that small constituencies and the d'Hondt rule have on breaking proportionality.

Rae and Ramírez. 1993. *El Sistema Electoral Español: Quince Años de Experiencia*, McGraw-Hill.
Urdanoz, 2008, "El maquiavélico sistema electoral español" *El País*, 16 February, 35.
Santoyola, 2011. "III Jornadas de derecho parlamentario. La reforma de la ley orgánica del régimen electoral general", *Cuadernos Manuel Giménez Abab*, 1, 95-98.

Spanish Constituencies



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 gained similar among 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:

(1) Calculate the $K \times M$ matrix of quotients p_k/d_j, $k=1,\ldots,K$, $j=1,\ldots, M$, with $d_j=j$.

(2) Select the M largest quotients and give the corresponding parties a seat for each of their largest quotients.

The d'Hondt rule

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.

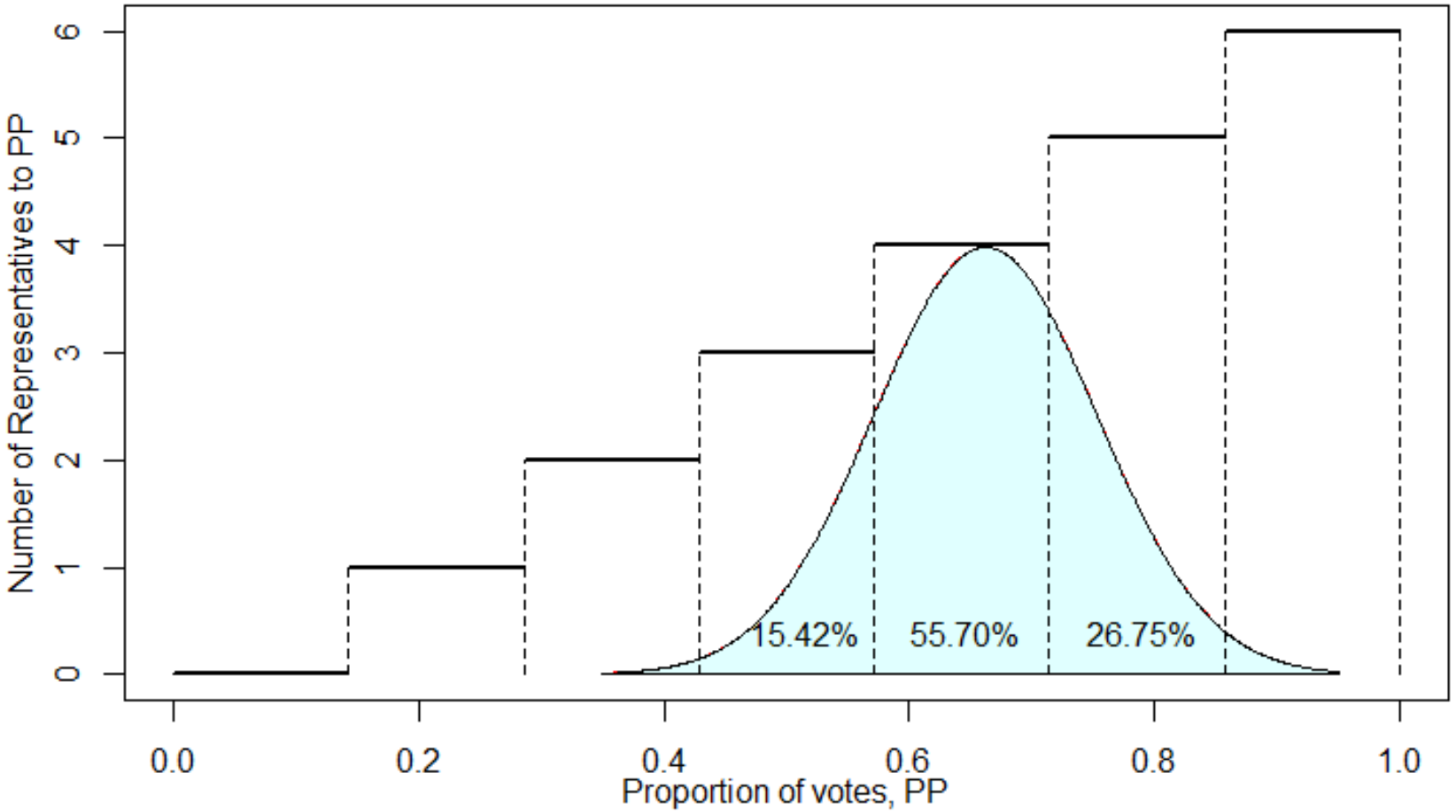
Example: 7 seats distributed among four parties (A, B, C, and D) receiving 45%, 32%, 15%, and 8% of votes produces 4, 2, 1, and 0 seats.

	Seats						
Party	1	2	3	4	5	6	7
A	**45.0**[1]	**22.5**[3]	**15.0**[5]	**11.3**[7]	9.0	7.5	6.4
B	**32.0**[2]	**16.0**[4]	10.7	8.0	6.4	5.3	4.6
C	**14.0**[6]	7.5	4.7	3.5	2.8	2.3	2.0
D	8.0	4.0	2.7	2.0	1.6	1.3	1.1

Can election poll district proportions be used directly to generate seats forecasts?

- In election polls, at most a couple of thousand electors are typically interviewed.
  (Average 2008-2011: 1,480. Last week 2011 election: 4,726)
- Predictions of share of votes are only reliable in (national or regional) aggregated terms.
- But representatives are elected in subnational constituencies: provinces.

Proportion estimates in districts show extreme variability



Constituency of Toledo. A *de facto* two party district, 6 seats in 2011 Election. National sample size: 2500. Sampling in ideal conditions. Normal approximation. 2011 General Election. Two party share of votes, PP: 68.22%; PS: 33.78%.

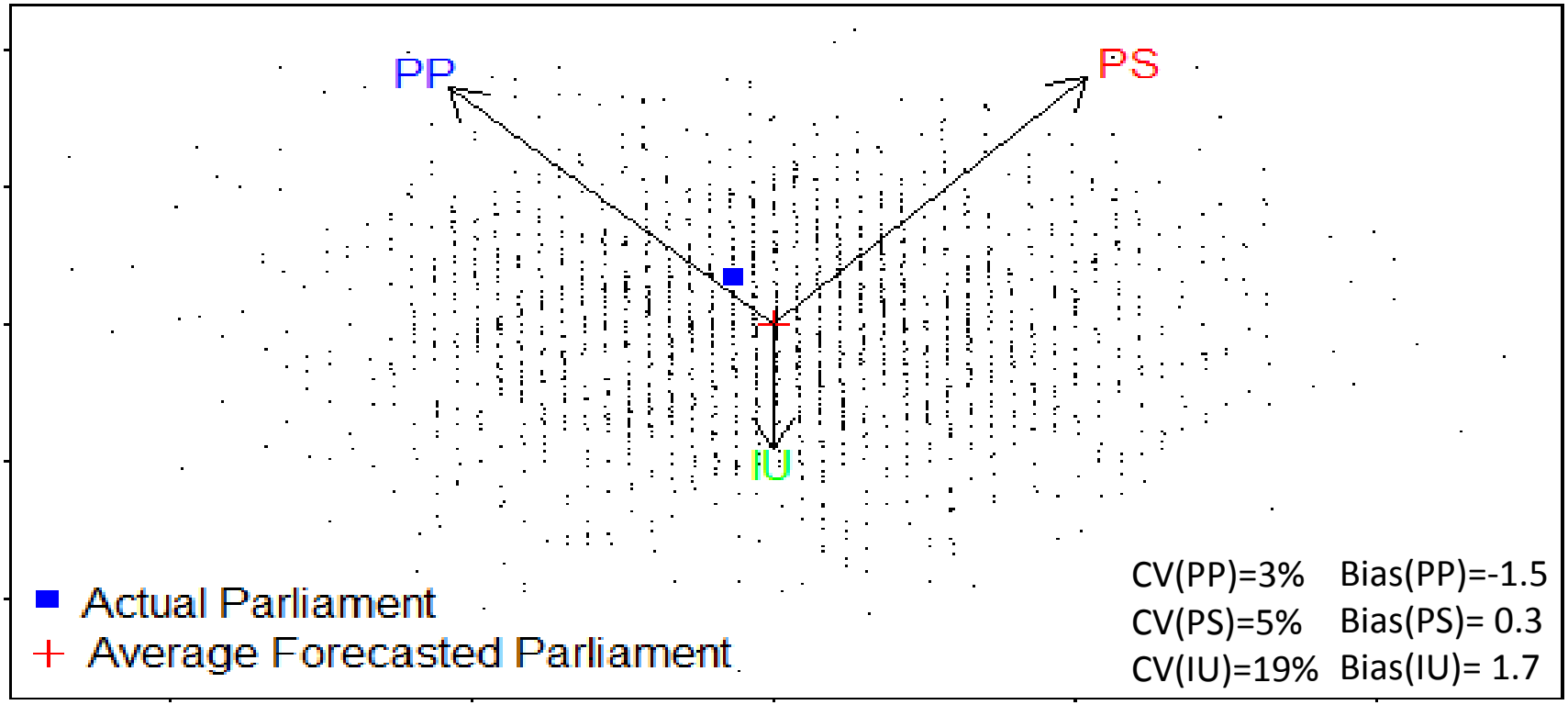
… and seat allocation produces aggregate bias

Although when there are many constituencies we might expect the district biases to cancel one another out and that the prediction for the whole Parliament would 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).

According to those authors, this may occur because (i) some locally important parties contend only in a few districts and due to (ii) several districts may have a similar bias as a consequence of general voting national patterns and small constituency sizes.

Delicado and Udina, 2001, "¿Cómo y cuánto fallan los sondeos electorales?" *Revista Española de Investigaciones Sociológicas,* 96, 123-150.
Udina and Delicado, 2005, "Estimating parliamentary composition through electoral polls" *Journal of the Royal Statistical Society-A,* 168, 387-399.

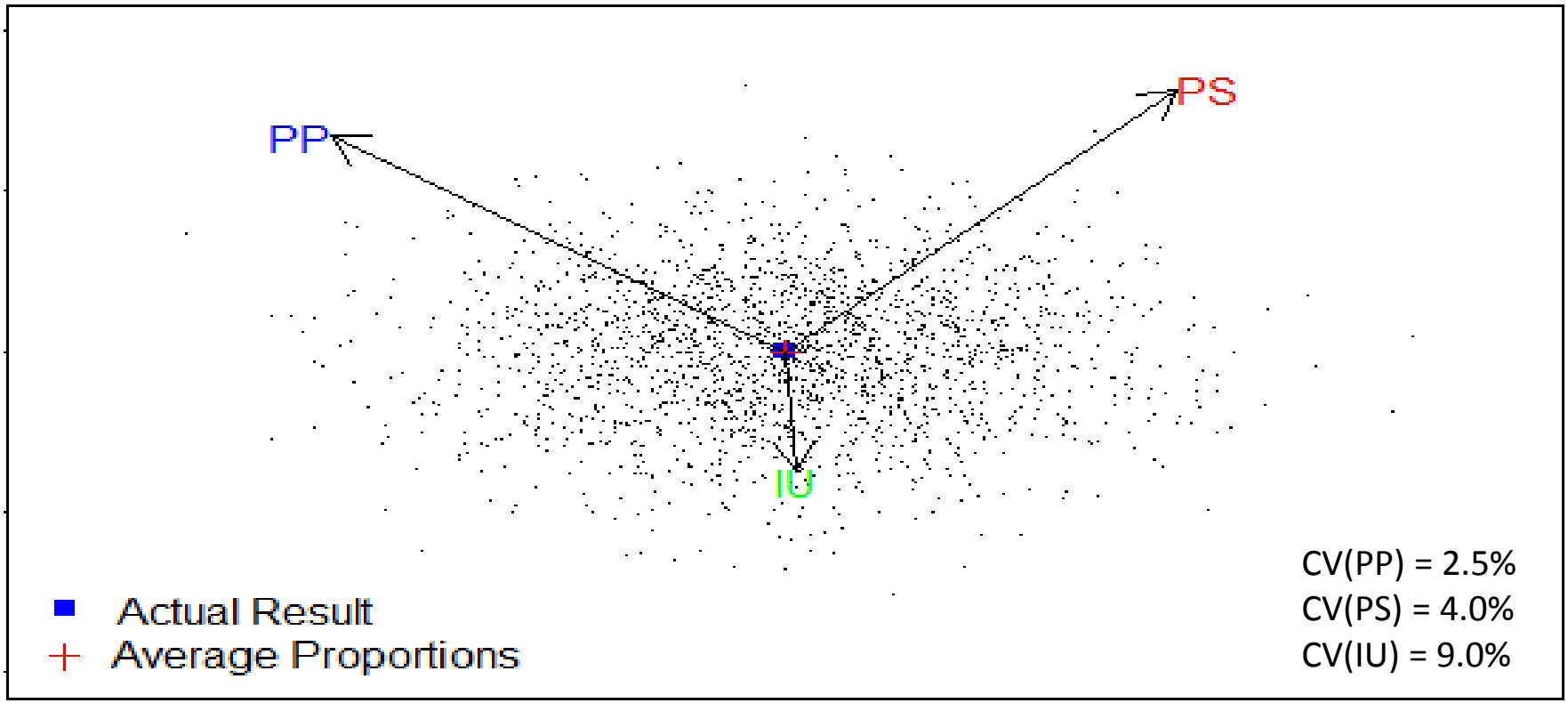
Poll Variability & Bias: Parliament Forecasts



The points in the scatterplot represent 2000 Parliaments projected on the plane defined for the two main principal components obtained from 2000 simulated polls of size 2500 using the 2011 Spanish General election final results. The two-components captured variance is 44.6%.

The arrows, starting from the average point of forecasted Parliaments, represent the directions favouring the three main parties. The blue square marks the position of the real Parliament and shows visually that there is a significant bias in the estimation of Parliamentary composition.

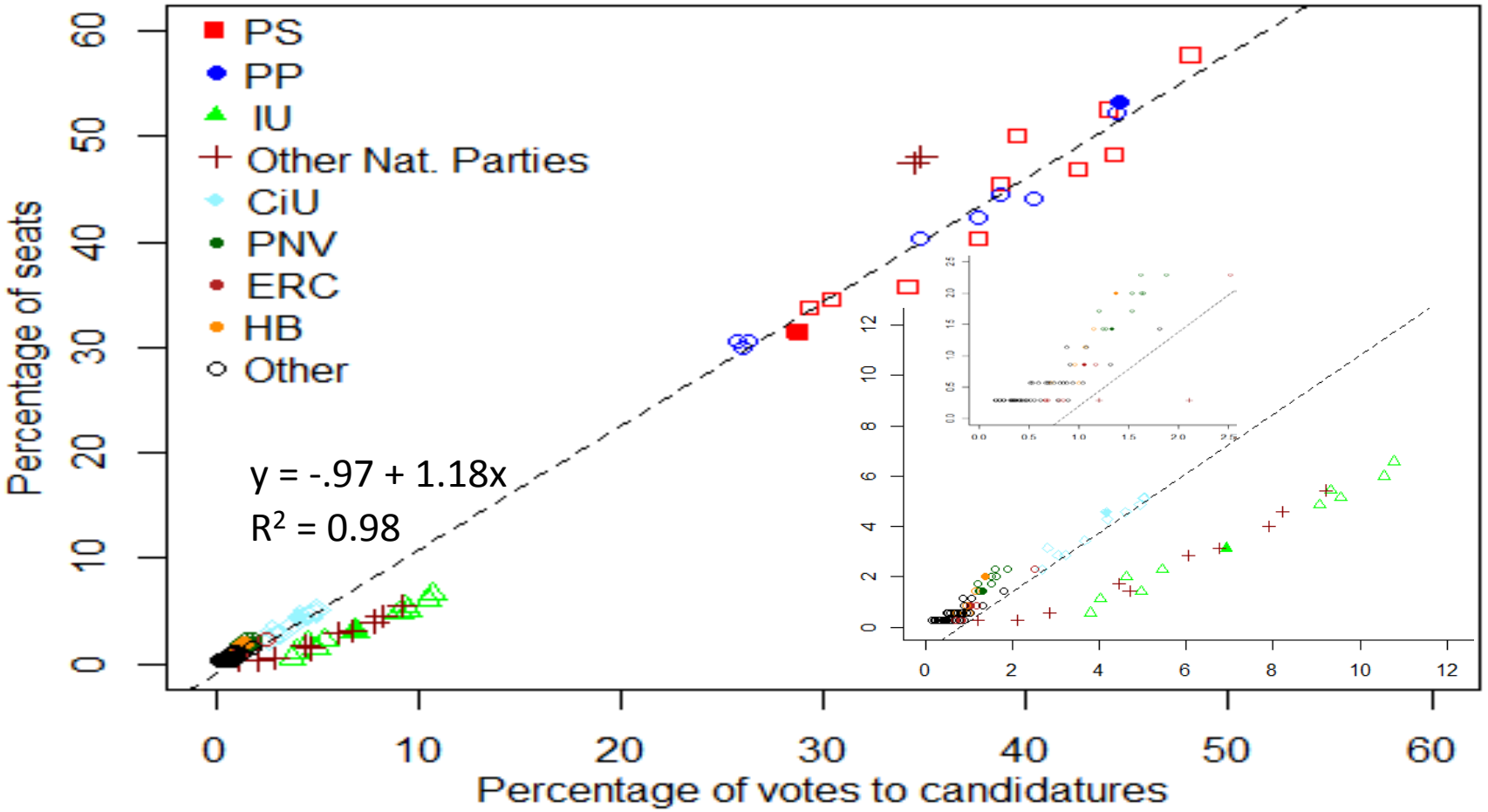
Poll Proportions: Less Variability & No Bias



The points in the scatterplot represent 2000 proportion polls projected on the plane defined for the two main principal components obtained from 2000 simulated polls of size 2500 using the 2011 Spanish General election final results. The two-components captured variance is 42.7%.

The arrows, starting from the average point, represent the directions favouring the three main parties. The blue square marks the position of the proportions and shows visually that there is no bias in the estimation of proportions.

Looking for historical relationships between votes and seats (1977-2011)



Modelling the votes-seats relationship

A strong linear relationship exists between the proportions of votes and seats a party gains.

Despite the strong global relationship (R^2=0.98), it seems that a single equation for each party can yield better results.

For small national parties, the linear relationship seems to work after 4% of votes, a logit transformation maybe could be useful for the whole range.

For small regional and for *sporadic* parties, as well as 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.

Model-I

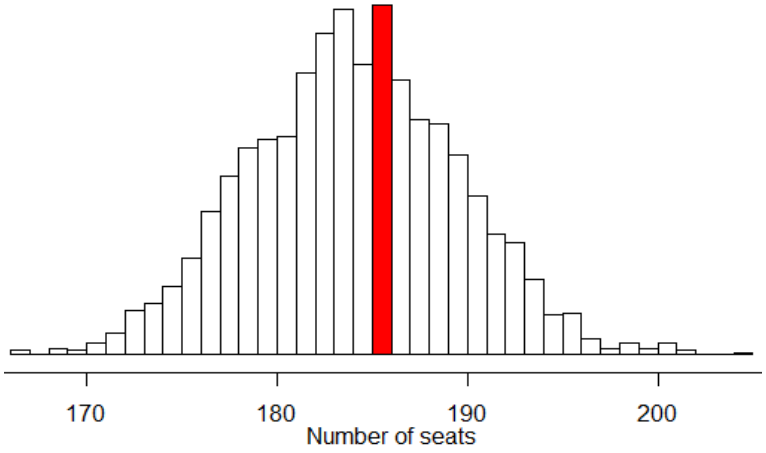
1. Fit univariate linear equations for PP and PS.
2. Fit a univariate linear equation for small national parties (range 4%-12% of national votes).
3. Fit a univariate linear equation for Catalonian regional parties.
4. Fit a univariate linear equation for regional parties from The Basque Country.
5. Fit a univariate linear equation for other regional parties.
6. Round to the nearest integer the seats forecasts (except for point 5 predictions that are round to zero).
7. And, given that the Spanish system favours the party obtaining the great support, to assign to the biggest party the difference to 350 seats.

Comparing Poll & Model-I Forecasts

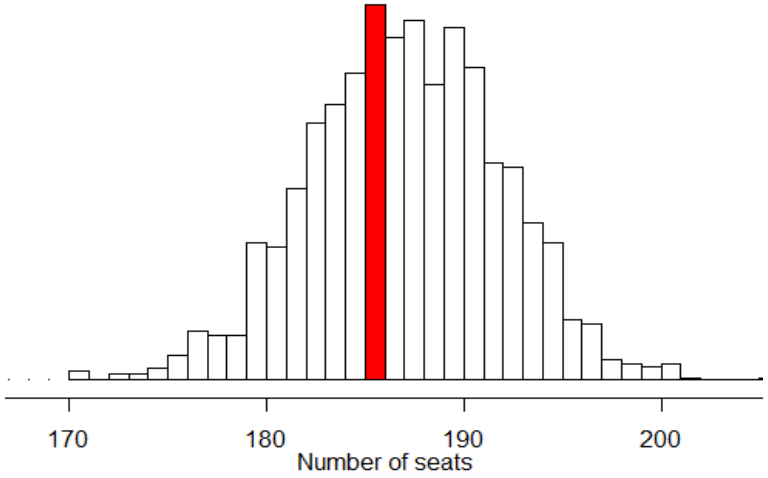
	Actual	Direct Poll Forecasts				Model I Forecasts			
Party	**Results**	**Average**	**Bias**	**Variance**	**MSE**	**Average**	**Bias**	**Variance**	**MSE**
PP	186	184.5	-1.5	30.2	**32.4**	187.5	1.5	21.3	**23.6**
PS	110	110.3	0.3	28.7	**28.8**	111.1	1.1	23.8	**24.9**
IU	11	12.7	1.7	6.1	**8.9**	11.9	0.9	3.0	**3.8**
UPyD	5	6.5	1.5	3.2	**5.6**	6.0	1.0	2.2	**3.2**
CiU	16	16.7	0.7	4.4	**4.9**	14.4	-1.6	3.7	**6.2**
Esquerra	3	2.6	-0.4	1.3	**1.4**	2.4	-0.6	1.1	**1.5**
PNV	5	5.2	0.2	1.6	**1.6**	5.4	0.4	1.6	**1.8**
Amaiur	7	5.3	-1.7	2.1	**4.9**	5.7	-1.3	1.7	**3.5**
BNG	2	1.5	-0.5	0.9	**1.1**	1.6	-0.4	0.4	**0.5**
NA.BAI	1	0.5	-0.6	0.3	**0.6**	0.1	-0.9	0.1	**0.9**
CC	2	2.0	0.0	1.0	**1.0**	1.2	-0.8	0.3	**0.9**
Q	1	0.7	-0.3	0.4	**0.5**	1.0	0.0	0.3	**0.3**
FAC	1	1.1	0.1	0.4	**0.5**	0.8	-0.2	0.2	**0.3**
PA	0	0.1	0.1	0.1	**0.1**	0.5	0.5	0.3	**0.6**
PRC	0	0.4	0.4	0.3	**0.4**	0.1	0.1	0.1	**0.1**

Results based on 2000 simulated polls of size 2500 (with a fixed allocation of 15 units per district and the remaining 1750 units allocated proportional to district populations) using the 2011 Spanish General election final results and under random simple sample in each constituency.

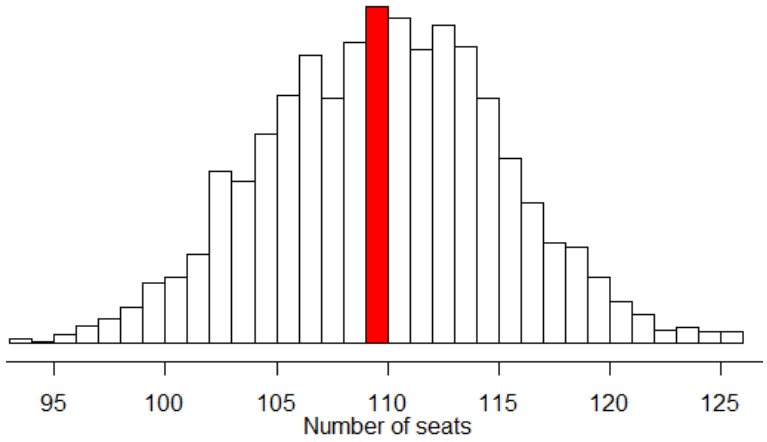
Comparing Poll & Model-I Forecasts



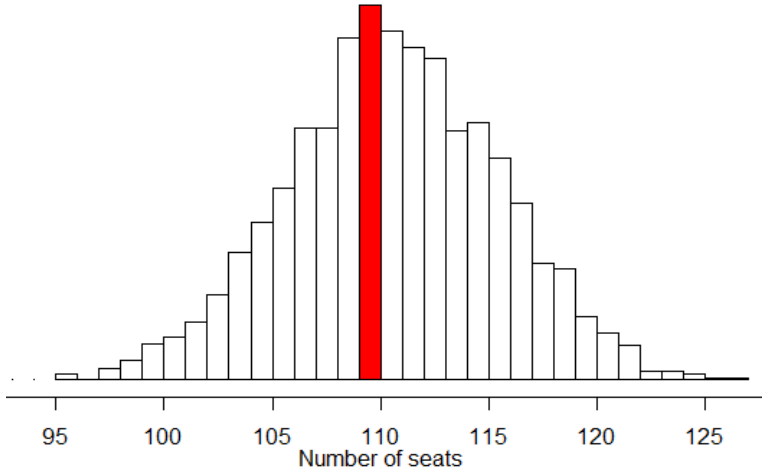
PP Poll Forecasts



PP Model-I Forecasts



PS Poll Forecasts



PS Model-I Forecasts

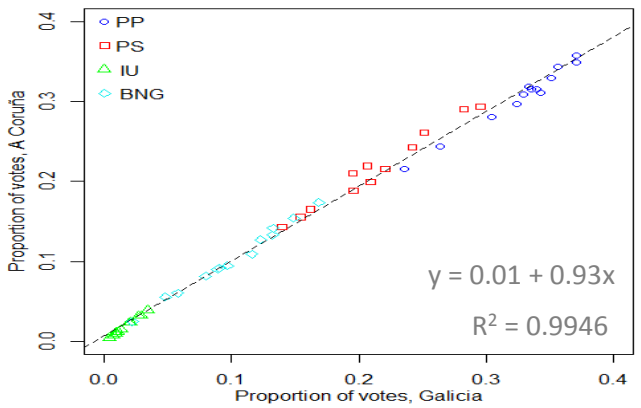
Looking for other relationships

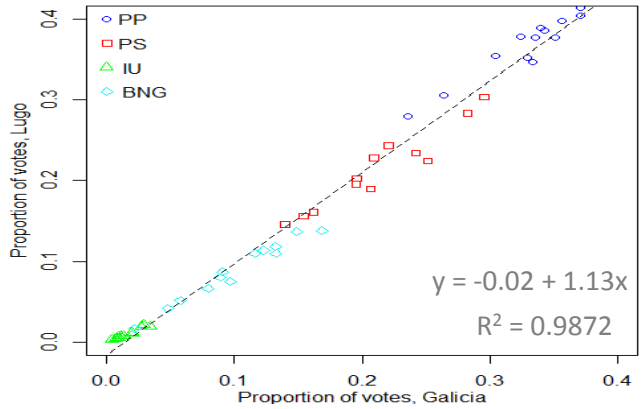
Despite the improvement that entails using the votes-seats model (the total MSE is reduced by 30%), Model I is quite sensible and translates almost all the variability of the poll estimates.

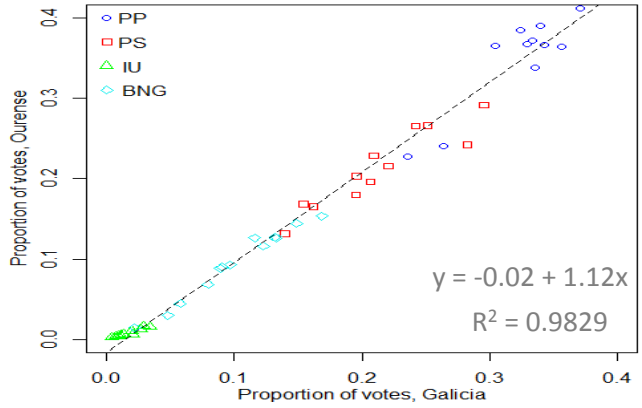
And although (i) there is room to improve Model I and (ii) poll variability could be significantly reduced by using ratio estimators, post-stratification techniques or superpopulation models (Mitofsky and Murray, 2002; Mitofsky, 2003, Pavía and Larraz, 2012), in what follows we explore the impact of using global-constituencies votes relationships.

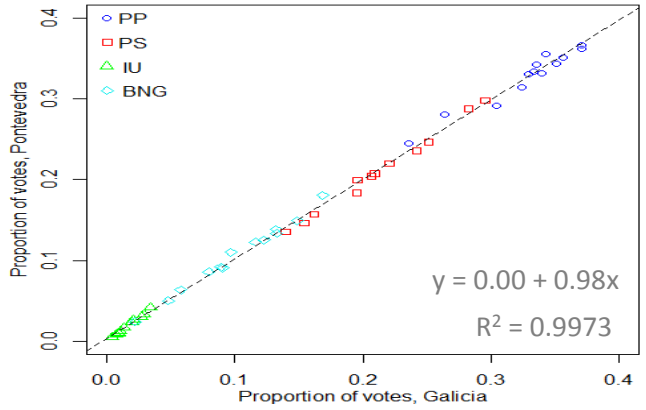
Mitofsky and Murray, 2002, "Election Night Estimation", *Journal of Official Statistics*, 18, 165-179.
Mitofsky, 2003, "Voter News Service after the Fall", *Public Opinion Quarterly*, 67, 45-58.
Pavía and Larraz, 2012, "Nonresponse Bias and Superpopulation Models in Electoral Polls" *Revista Española de Investigaciones Sociológicas,* 137, 237-264.

Global-Constituency vote relationships Constituencies of Galicia (1989-2011) Regional and General elections









Model-II

1. Fit univariate linear models per each party and constituency between the total proportion of votes and the corresponding party-constituency proportion of votes.
2. Use the proportion estimates obtained for each party in each constituency in (1) to allocate seats in the corresponding constituency.
3. Use for *new* parties the relationship of the party ideologically closer.

Comparing Poll & Model-II Forecasts for 2012 Galician regional elections

	Actual	Sampling Forecasts				Model II Forecasts			
Party	Results	Average	Bias	Variance	MSE	Average	Bias	Variance	MSE
PP	41	40.8	-0.2	4.8	**4.8**	41.0	0.0	4.8	**4.8**
PS	18	17.5	-0.5	3.6	**3.9**	17.7	-0.3	2.6	**2.6**
EU-ANOVA	9	9.8	0.8	2.6	**3.3**	9.3	0.3	1.9	**1.9**
BNG	7	6.9	-0.1	2.1	**2.1**	7.0	0.0	1.8	**1.8**

Results based on 2000 simulated polls of size 800 (with a fixed allocation of 100 units per district and the remaining 400 units allocated proportional to district populations) using the 2012 Galician Parliament regional election final results and under random simple sample in each constituency. Seats: A Coruña 24, Lugo 15, Ourense 14 and Pontevedra 22 . In each constituency a threshold of 5% of valid votes.

In this case, despite the larger number of seats allocated per constituency, which makes model improvements less evident, an improvement in both bias and variance is obtained, with a reduction of total expected Mean Squared Error of 21%.

Conclusions & Further Remarks

- Two model-based approaches to translate poll proportions into seats has been proposed in this research within the Spanish electoral context.
- The suggested approaches significantly reduce the sampling variability associated with polls, but their forecasts still show too much volatility.
- Despite the simpler relationships that link votes and seats in PR systems, the higher sensitivity that seat share exhibits in relation to small variations in share of votes makes of this problem a difficult and interesting challenge.

Conclusions & Further Remarks

- The analysis performed in this presentation show that this avenue of research looks promising and that there is still enough room to improve.
- Within the proposals presented here, both models deal with linear univariate relationships. They could be extended to multivariate versions (using, for example, SURE models) in which historical correlations can be taken into account.

Conclusions & Further Remarks

- Other possible extensions that also deserve attention would include (i) to consider the spatial dimension of the data, (ii) the use of small area estimation models in which shrinkage constituency estimates were obtained combining model-based predictions and constituency poll estimates, or (ii) to fit the relationships through multinomial models.