

# Are Election Polls Right? Analysis of the Accuracy of Election Polls Predictions Published in the Press

*¿Aciertan los sondeos electorales? Análisis sobre la bondad de predicción de los sondeos electorales publicados en la prensa*

**Adrià Caballé, Pere Grima and Lluís Marco-Almagro**

## Key words

Surveys • Elections  
• Fieldwork • Electoral Behaviour • Mass Media • Distribution of Parliamentary Seats  
• Research Design

## Palabras clave

Encuestas • Elecciones  
• Trabajo de campo  
• Conducta electoral  
• Medios de comunicación  
• Distribución de escaños • Diseño de la investigación.

## Abstract

This paper analyses the predictions of the election results for the Spanish Parliament published in the main Spanish newspapers. To gauge their degree of accuracy, the distance between the predictions and the actual results was defined and later assessed by comparing it with that of the so-called 'perfect poll'. This 'perfect poll' was generated by a simulation using the proportions of votes actually obtained by each party. Other issues related to the published polls are also analysed, such as the relationship between sample size and estimation quality, and between the error made (by excess or defect) for a party and the ideological affinity between that party and the newspaper publishing the survey.

## Resumen

Este artículo analiza las predicciones sobre los resultados de las elecciones al Congreso de los Diputados de España publicadas en los principales medios de la prensa española. Para comparar su grado de acierto se define una distancia entre la previsión y los resultados reales, y se valora su magnitud comparándola con la que presenta lo que se ha denominado "sondeo perfecto", realizado por simulación a partir de las proporciones realmente obtenidas por cada partido. También se analizan otros aspectos relacionados con los sondeos publicados, como la relación entre el tamaño de muestra y la calidad de la estimación realizada, o entre el error que se comete (por exceso o por defecto) con un partido y la afinidad ideológica entre dicho partido y el medio que publica la encuesta.

## Citation

Caballé, Adrià; Grima, Pere and Marco-Almagro, Lluís (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.  
(<http://dx.doi.org/10.5477/cis/reis.143.25>)

**Adrià Caballé:** Universidad Politécnica de Cataluña | [adriacaballe@yahoo.es](mailto:adriacaballe@yahoo.es)

**Pere Grima:** Universidad Politécnica de Cataluña | [pere.grima@upc.edu](mailto:pere.grima@upc.edu)

**Lluís Marco-Almagro:** Universidad Politécnica de Cataluña | [lluis.marco@upc.edu](mailto:lluis.marco@upc.edu)

## INTRODUCTION<sup>1</sup>

A certain degree of controversy has always existed regarding the prediction ability of election polls. On the one hand, they are typically referred to with a tone of skepticism (“they are only polls”) particularly by those who do not benefit from their predictions. Yet on the other hand, they have become a common feature of today’s media, serving as guides for commentaries and interpretation of the current political reality.

Two types of studies may be looked at in regards to voter intention: those conducted throughout the length of the legislature, often having titles such as “if the elections were held today ...”, and those that are published one week prior to the elections. In the first case it is impossible to know if the predicted results have a reasonable level of accuracy or not, as is typical of studies based on sample analysis (such as TV audience studies, for example). However, for polls that are published one week prior to the elections, the “true” outcome is in fact ultimately known, and therefore these predictions may be evaluated objectively.

In addition to this unique characteristic, as well as the fact that many individuals relate statistical possibilities with the accurate results of these prediction types, it is clear that election polls are quite complex in nature, at times extending beyond the capabilities of statistical techniques and may be summarized by the following points:

- Difficulty in obtaining a representative sample: This point is not specific to election polls; any prediction based on a sample study requires that the sample be representative and achieving this is neither easy nor cheap. Survey documentation from many election polls indicates that the surveys tend to be conducted by telephone, and in these cases, such factors as the time of the call, the individual with whom the caller asks to speak with or who is actually spoken to (in substitution of unwilling participants), are of great importance. The knowledge and experience of companies specializing in achieving representative samples at a competitive cost is a key to success. Díaz de Rada (2007) offers a list of inherent difficulties that all face-to-face surveys may encounter. Surveys conducted online also face their own set of difficulties (Pavía et al., 2011).
- The intention to change votes: Election polls that are published one week prior to the elections may be based on surveys conducted various days, and even up to a week prior to their publication. Therefore, in addition to extrapolations made from the sample to the population (with sample statistic theory dealing with this) generalizations must also be made from the dates on which the surveys were conducted to the actual election day, and during this time period, any number of events may occur affecting the final decision made by both decided and undecided voters.
- Allocating undecided votes: The problem here lies in the fact that those individuals who do not respond to polls are not a random sample of the population. If the reasons for not responding were statistically independent of voting preferences then this could be easily resolved, statistically speaking. However, it has been determined that this is not the case and the problem is therefore quite complicated, as Bouza (1998) explains.

---

<sup>1</sup> Our interest in this subject arose from discussions held with professor Roberto Behar, of the Universidad del Valle en Cali (Colombia) and professor Pedro Delicado of the Universitat Politècnica de Catalunya (UPC), to whom we wish to offer our most sincere gratitude. We also wish to thank professor Josep-Antón Sánchez-Espigares of the UPC who offered assistance in statistical and computational aspects of the study. To Victor Peña, Master’s student of Statistics and Operational Research at the UPC, for his detailed reading of the original version, allowing us to correct any errors. And finally, to the two anonymous journal evaluators whose detailed analysis of our work and suggestions have allowed for its considerable improvement.

- Lack of response or lack of sincerity: Some individuals prefer not to express their political preferences (as opposed to *undecided*, they may be referred to as “guardedly decided”). This reservation to express the intent to vote is more commonly found in those who plan to vote for determined political parties. Pavía and Larraz (2012) have extensively analyzed this problem which they define as “the primary source of error in electoral predictions in Spain”, suggesting strategies to address it.
- Allocating the number of obtained seats based on the percentage of votes: Statistical laws may serve to estimate the proportion of votes, but the truly relevant information is the number of seats obtained. If, in a determined constituency, there are 5 seats at stake and it is possible to predict with 95% confidence that a determined party will obtain 32% of the votes with a 3% margin of error, the problem is that if the party obtains 31% they will receive only one seat yet if they obtain 33% they will receive two. This is a major difference, but available information does not permit the choice of one option or the other. Another difficulty encountered is the need to reach 3% of the votes in order to enter into the allocation of seats; some parties may almost reach this percentage, but it may be impossible to know if they attain it or not, with possible repercussions in the overall distribution. Finally, application of the D’Hondt method introduces certain prediction problems: while the prediction of the proportion of votes is unbiased (the average of the obtained percentages, if many polls are conducted, will coincide with the actual percentages), the number of seats obtained applying the D’Hondt method offers a skewed prediction, as described by Delicado and Udina (2001).

The main objective of this article is to respond to the question proposed in the title, offering arguments based on data analysis.

Naturally, the term “accurate” should not be taken at face value (as it is not possible to make truly “accurate” predictions), but rather, in the sense of whether or not the predictions can be considered reasonable, taking in to account the mathematical margin of error intrinsic to any statistical estimation<sup>2</sup>.

## ANALYZED POLLS

Analysis was conducted of polls conducted for elections to the Spanish Congress of Deputies as published in major Spanish newspapers: *El País*, *El Mundo*, *ABC*, *La Vanguardia* and *El Periódico*<sup>3</sup>. These polls covered the election periods from the 22<sup>nd</sup> of June of 1986 until the 20<sup>th</sup> of November of 2011, as no polls were conducted by the relevant media for the first three elections taking place after the country’s return to democracy (15<sup>th</sup> of June of 1977, 1<sup>st</sup> of March of 1979 and 28<sup>th</sup> of October of 1982).

Annex 1 contains the results of the analyzed polls, ordering the parties in accordance with the number of seats obtained, as well as the date and the newspaper page in which they were published.

## MEASURING PREDICTION QUALITY

In order to evaluate and make comparisons between different polls, it is necessary to define a measure to quantify prediction quality.

<sup>2</sup> For terminology purposes, we should clarify that on the one hand there are those results obtained from application of statistical methods to the sample data (considered polling results *stricto sensu*) and on the other hand, there are those results that are published after being revised in accordance with the criteria of the company conducting the polling. Our analysis refers to published results, which always include adjustments made by the responsible companies. However, this is the only possible analysis, given that direct citizen responses from the sample are never published.

<sup>3</sup> Office of Measurement Justification (OJD): <http://www.ojd.es/OJD/> (Consulted on the 24<sup>th</sup> of November of 2011)

Among the various existing possibilities (see Manly, 1986 for examples) the following may be considered.

*Aggregate distance (AD)*

This is equal to the sum of the differences, in absolute value, between the predictions made and the values obtained for each party. That is,  $e_1, e_2, \dots, e_k$  are the seats obtained by  $k$  parties running in an electoral race and  $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_k$  are the predictions published in a specific poll. Thus, the *aggregate distance* between the predictions and actual results is:

$$DA = \sum_{i=1}^k |e_i - \hat{e}_i|$$

For example, the predictions published in *El Periódico* for the elections from the 22<sup>nd</sup> of June of 1986 and the actual results were as follows:

	PSOE	AP	CDS	CiU	IU	PNV	HB	EE	Others
<i>El Periódico</i>	180	98	30	17	5	7	4	1	8
Actual results	184	105	19	18	7	6	5	2	4
Difference (absolute value)	4	7	11	1	2	1	1	1	4

Therefore, the *AD* between the predictions made and the actual results is:

$$DA = 4+7+11+1+2+1+1+1+4 = 32$$

*Euclidean distance (ED)*

The results obtained by  $k$  parties may be considered to be a vector in a space of  $k$  dimensions and the *Euclidean distance* is the ordinary distance in this type of space, with its expression being a generalization of the Pythagorean Theorem:

$$ED = \sqrt{\sum_{i=1}^k (e_i - \hat{e}_i)^2}$$

*Standardized Euclidean Distance (SED)*

The error –in absolute value– in the prediction of the number of seats tends to be greater in those parties obtaining a larger number of representatives. Therefore, for example, it is more likely to commit an error of  $\pm 4$  representatives in the prediction of results for the PSOE in the 1986 elections, where they received 184 seats, than in the prediction made for the IU, receiving only 7, or of the EE, with 2 (in this latter case, only an underestimation error of 2 is possible).

This is due to the fact that the variability of the vote percentage estimator increases in a manner that increases said percentage until reaching its maximum at 50%. Therefore, it is reasonable to consider standardizing the difference between the estimated value and the actual value with a measure of prediction variability

that depends on the proportion of obtained votes for each party. Considering, for simplification purposes, that the variabilities are independent, the standardized Euclidean distance formula is as follows:

$$SED = \sqrt{\sum_{i=1}^k \left( \frac{e_i - \hat{e}_i}{\sigma_i} \right)^2}$$

with  $\sigma_i$  being the standard deviation of the estimator of the number of seats obtained for party  $i$ . If a party obtains a proportion of votes  $p_i$ , the variance of the estimator of said

proportion is in accordance with the formula:

$$V(\rho) = \frac{\rho_i(1-\rho_i)}{n}$$

However, the variance of the estimator of the number of seats is not so easily determined since the D'Hondt method destroys its mathematical properties (see Udina and Delicado, 2005) and must be estimated through simulation. Covariance between distinct parties receiving the so-called Mahalanobis distance may also be included, but as it is not of interest in this case, we feel that this is not appropriate.

Comparing these distances, it is clear that the simplest and easiest to interpret is the *aggregate distance*. The Euclidean distance produces an ordering of the predictions similar to that of the aggregated distance (the

## DISTANCE BETWEEN POLLS AND RESULTS

To calculate this difference, there are two initial difficulties which must be overcome: the use of intervals in making predictions and the fact that the "others" category does not always include the same number of parties.

The problem of intervals may be resolved by generating all of the possible result combinations and calculating the distance only for those adding up to the expected 350 representatives. The average of these distances, corresponding to demonstrable results, is the value assigned as the distance between the prediction and the actual result.

For example, the poll published by *El País* for the elections held in November of 2011 was as follows:

	PP	PSOE	CiU	IU-LV	AMAIUR	UPyD	EAJ-PNV	ERC	Others
<i>El País</i> Poll	192-196	110-113	14	11	4-5	2	5	3	5
Actual results	186	110	16	11	7	5	5	3	7

greater the difference in absolute value, the greater the squared difference), but in our case, having  $n$  parties, it is not as intuitive. The *standardized Euclidean distance* can be considered more exact as it values the produced errors related to probability, but it has the inconvenience of further penalizing the errors made regarding the smaller parties, when what is of the greatest interest here are the results with respect to the larger parties, which are those that have a greater influence on the new government. Therefore, it is considered that the *aggregated distance*, in addition to being clear and easily understood, can also easily adapt to our interests for this study and therefore is the one that we have selected to use.

In this case, for the PP, 5 possible results were presented, while 4 were presented for the PSOE, 2 for AMAIUR and 1 for the remainder, with the total number of combinations being:  $5 \times 4 \times 2 = 40$ , but only 9 of which may be considered, since only they add up to 350. The average of the distances corresponding to these 9 combinations of demonstrable results will be the distance assigned to this poll (in this case, equaling 19).

It may be seen that the only way to obtain a distance equaling zero is by correctly predicting the results for all of the parties with concrete estimates, not intervals, and that the proposed method penalizes wide intervals, since while they may include the actual result, they also include many inaccurate predictions that increases the distance.

With regards to the second difficulty (not all polls publishing predictions for the same number of parties), if the distance is always calculated using all of the poll estimates, those having the greatest amount of detail (including more parties) are penalized and this is not reasonable. The results of the poll published by *El Mundo* for the 2008 elections are presented below:

The prediction made for IU, BNG and CC offers some value to the distance between the polling and the actual results, as the poll values and those of the actual results for the IU and BNG do not coincide. However, if the-

way, in addition to applying homogenous criteria, it is possible to include all of the polls published in the considered period in the study (Table I).

These results are presented in Graph 1, showing that the predictions made for the 2008 elections were the best and that there was a very similar accuracy of predictions for all of the evaluated newspapers, with the 2004 predictions (following the 11M attacks) being the worst. It is also found that in some elections, such as those of 1996, some newspapers published polls that were much more accurate than others.

	PSOE	PP	CiU	EAJ-PNV	ERC	IU	BNG	CC	Others
<i>El Mundo</i> Poll	157-171	148-161	9-11	7	5-6	4	0	2	2
Actual results	169	154	10	6	3	2	2	2	2

se three parties had been grouped into a new category "Others" (including the estimate "2" that already exists for this category), the prediction would coincide with the actual results (prediction= actual results= 8) and there would be a smaller distance.

In order to remove the advantage resulting for those polls having larger "Others" groupings, and in order to attain comparable results, distances have been calculated with predictions made for the 5 parties obtaining the greatest number of seats, grouping the remainder in the "Others" category. In this

The question that now arises is whether or not the observed distances are large or small with respect to that which is expected. To answer this question, we have simulated what we refer to as "perfect polls".

## A REFERENCE DISTANCE: THE PERFECT POLL

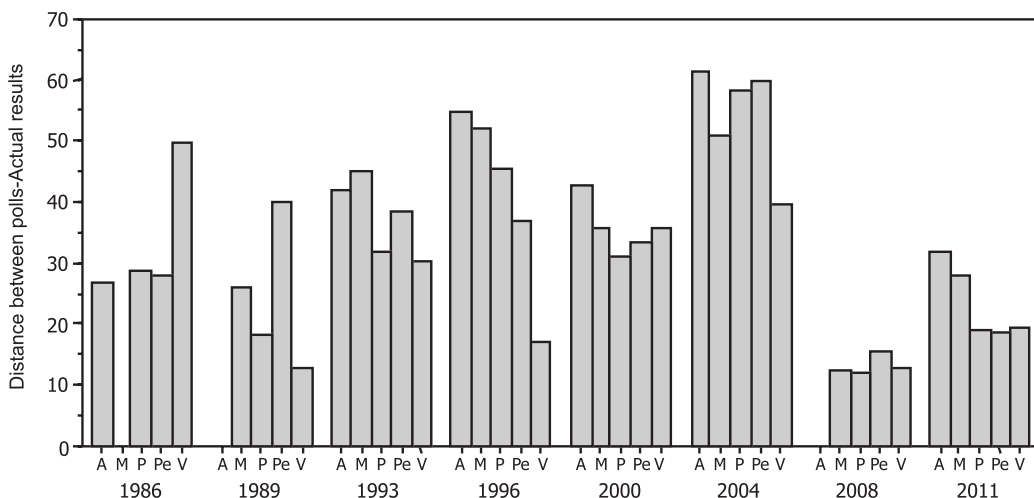
By *perfect poll* we refer to a poll which has an appropriate sample size and in which all of those surveyed tell the truth: if they will vote or abstain from voting and, in the case in which

**Table 1:** Distances between the polls and the final results of the general elections

	1986	1989	1993	1996	2000	2004	2008	2011
<i>ABC</i>	26.8	-	42.1	54.8	42.8	61.6	-	32
<i>El Mundo</i>	-	26	45	52	36	51	12.3	28
<i>El País</i>	28.8	18.2	32	45.7	31	58.2	12	19
<i>El Periódico</i>	28	39.9	38.2	36.8	33.4	59.8	15.5	18.6
<i>La Vanguardia</i>	49.7	13	30.54	17.2	36	39.8	13.1	19.4



**Graph 1:** Graphic representation of the distances between the polls and the actual results of the general elections from the indicated years (A: ABC, M: El Mundo, P: El País, Pe: El Periódico, V: La Vanguardia)



they will vote, for whom they will do so. The largest sample size appearing in the survey documentation of the published polls contains 15,600 individuals in a poll conducted by IP-SOS-EcoConsulting for the ABC in the 2000 elections. An appropriate sample size is understood to be large yet realistic, thus a sample of 20,000 individuals has been used for the perfect polls. For the year 2000 and following election, each constituency has had a total sample proportion assigned to it, equaling that used by the CIS in their pre-electoral survey<sup>4</sup>. For elections held prior to 2000, sample sizes were not explicitly published for each constituency and for these elections, the proportions from the year 2000 have been used.

After assigning a sample size to each constituency, a multinomial distribution was used—with the actual proportions obtained by each party—to simulate the results that would occur with a sample of this size in which all of those

surveyed always told the truth. Table 2 shows the actual results (proportion of votes and number of seats) for the province of Barcelona in the 2011 elections and the estimates obtained from the “perfect poll”. The votes in the sample are those obtained by simulation; total votes are an extrapolation of the total number of valid votes in the proportions obtained by each party in the sample, and number of seats is determined by applying the D’Hondt method to the total of the obtained votes.

By adding up the seats obtained for each constituency, an estimate of the general results provided for the perfect poll is attained. The corresponding distance of this estimate represents an inevitable error that cannot be eliminated despite conducting the sampling process perfectly, and without any of the inherent difficulties found in the electoral polls as commented earlier.

On the other hand, as the results of the simulation are influenced by chance, upon repetition of this process, the same values will not always be found. In order to quantify the distance corresponding to the perfect poll, it is repeated 10,000 times and the distance is measured for each of the 10,000 obtained results. The average of these values

<sup>4</sup> The poll’s survey documentation includes the total designed sample size and the size of each constituency. From this information, proportions are determined. Survey documentation for the year 2000 may be found at: [http://www.cis.es/cis/opencms/-Archivos/Marginales/2380\\_2399/2382/Ft2382.pdf](http://www.cis.es/cis/opencms/-Archivos/Marginales/2380_2399/2382/Ft2382.pdf) (Consulted on the 15th of December of 2011)

**Table 2:** Results prediction for the 2011 general elections in the Barcelona constituency using the "perfect poll"

		Actual results		Simulation estimates		
		% Votes	Seats	Total votes <sup>(1)</sup>	Total votes <sup>(2)</sup>	Seats <sup>(3)</sup>
Abstention		29.35	-	286	-	-
Percentage of the total valid votes	PSC-PSOE	27.78	10	175	498,318	9
	CIU	27.15	9	179	509,708	10
	PP	20.95	7	146	415,739	8
	ICV-EUiA	9.07	3	52	148,072	2
	ERC-RI.cat	6.48	2	46	130,986	2
	PxC	2.02	0	9	25,628	0
	Eb	1.53	0	9	25,628	0
	Others/blank	5.02	-	29	82,578	-

(1) The sample size for the Barcelona province should include 931 individuals, in accordance with the proportion applied by the CIS in the pre-electoral study of the 2011 elections (0.04656).

(2) Obtained sample proportion, extrapolated to the total number of valid votes.

(3) Application of the D'Hondt method to the total votes.

is considered to be the distance between the perfect poll and the final result<sup>5</sup>.

In Graph 2 the distances for the published polls have been again shown, but with the addition of a dark horizontal line to represent the value corresponding to the average distance represented by the perfect poll for these elections. The percentages of 97.5% and 2.5% have been determined as the values between which the average distance of the perfect poll is found in 95% of the cases. These limits are marked with dashed lines.

It is seen that in the 2008 elections, all of the published polls have a distance from the actual result that is within the interval offered by the perfect poll in 95% of the cases. In other elections, some of the polls are within the limits (1989) or quite close (1996 and 2011), while the others (1986, 1993, 2000 and 2004) have greater distances.

Graph 3 shows the distribution of the distances obtained in the 10,000 simulations of the perfect poll for the 2011 elections. In this same graph, the distances corresponding to the analyzed polls are presented.

The distribution of the distances obtained from the perfect poll simulations has a similar form in all of the elections-- slightly asymmetric with a tail that leans to the right. This asymmetry is more accentuated when the average value of the distance is smaller, since in these cases, the boundary having a distance equal to zero is more influential. Annex 2 presents the distribution of the distances obtained for the perfect poll for the eight examined elections.

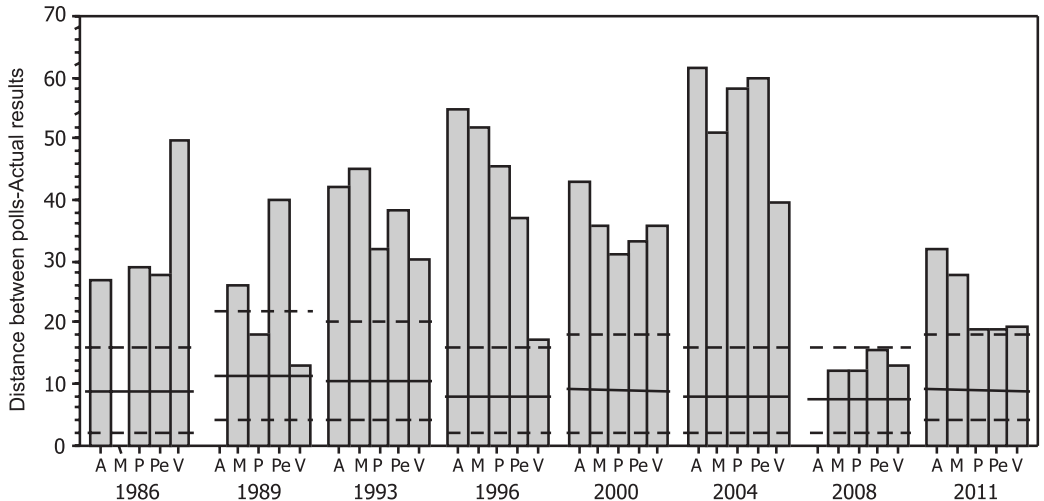
## PREDICTION ERROR DIRECTION AND POLITICAL LEANINGS OF THE PUBLISHING NEWSPAPER

It has been commented that newspapers tend to be optimistic in their predictions regarding the parties with which they share

<sup>5</sup> Results generation for the perfect poll and the calculation of distances has been made using programs written with the R Project for Statistical Computing.



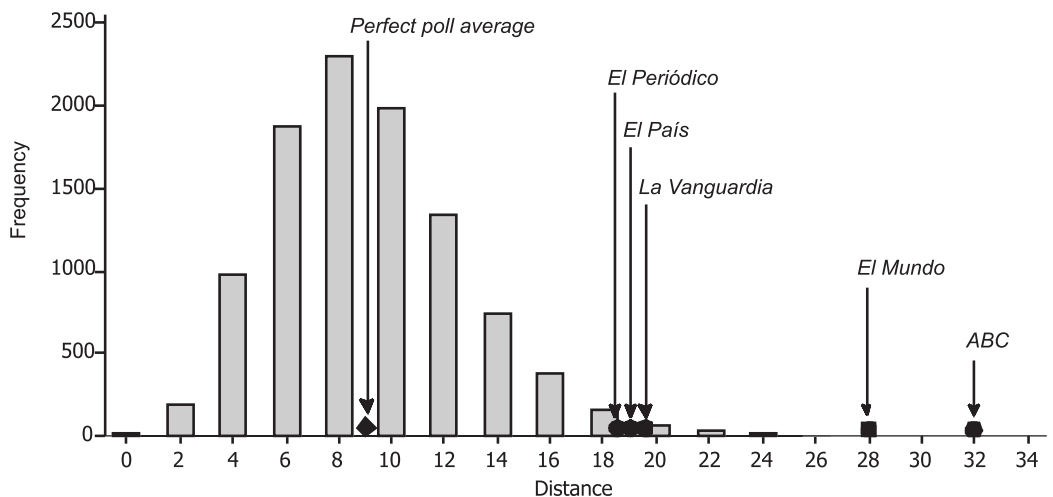
**Graph 2:** *Reproduction of Graph 1 with the addition of a dark horizontal line in each election, representing the average distance obtained from the perfect poll and including dashed lines that mark the area where this distance is found in 95% of the cases.*



ideological affinities and pessimistic regarding the contrary. In order to analyze whether data supports this possibility, those error committed in estimating the number of seats has been defined as the difference between the mean value of the prediction interval and

the actual value. For example, the prediction published by the ABC newspaper for the PSOE in the 1986 elections had an interval of 167–180 representatives and the actual value was 184, therefore, the error committed was  $(167+180)/2 - 184 = -10,5$ . It should be noted

**Graph 3:** *Distribution of the distances obtained in the perfect poll for the 2011 elections and values corresponding to the analyzed polls*



that this manner of defining error is equivalent to the mean of the differences between each value included in the interval and the actual value.

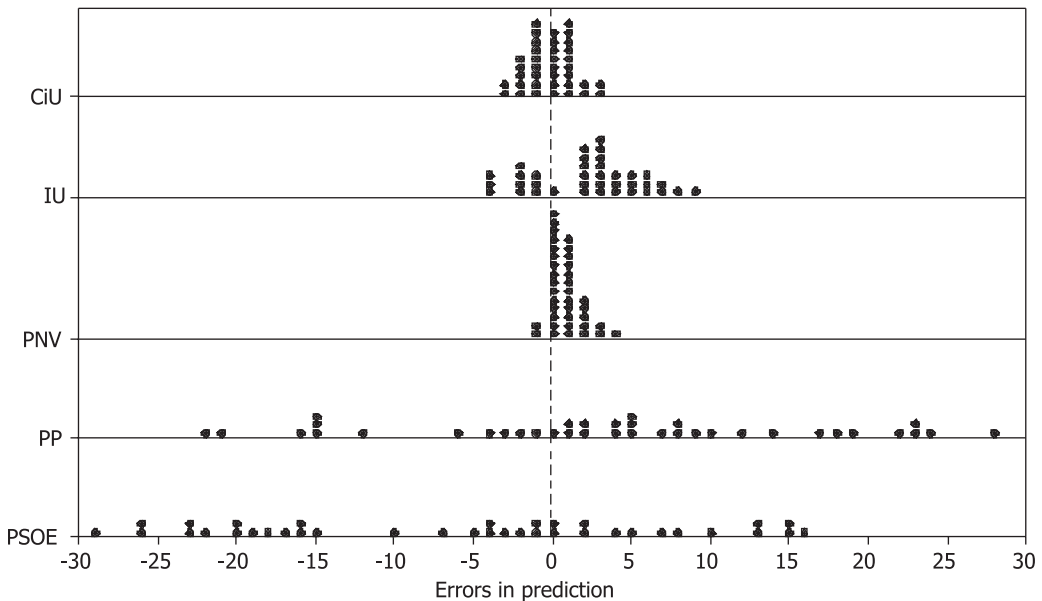
Graph 4 shows the errors committed for each party, grouping elections and newspapers (Annex 3 offers a detailed table including all of the data). It is evident that the errors are greater for those parties obtaining a larger number of seats (PSOE and PP) as predicted by statistical theory. For the CiU, the estimates were always quite good, whereas for the IU, PNV and PP, errors tended to be more in excess than in deficit, while for the PSOE there was a greater frequency of negative errors.

Graph 5 presents the average errors, upon separating the estimates based on newspaper but maintaining their grouping based on all of the elections (the averages are for the 8 evaluated elections). It is seen that the accuracy of the predictions made for the CiU and the PNV as well as the overesti-

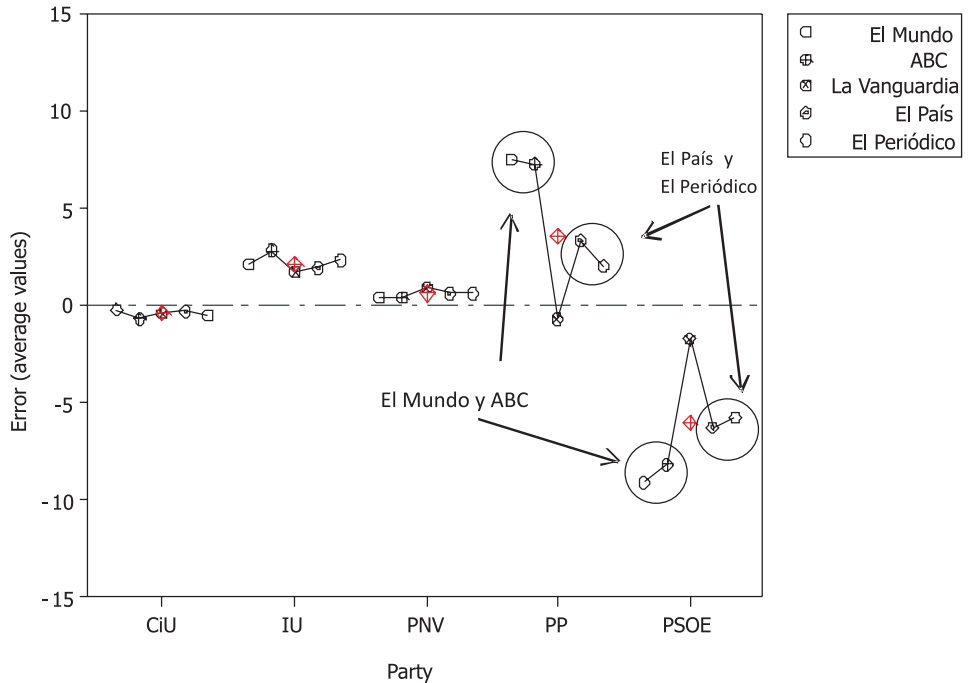
mation of predictions made regarding the IU occurred in similarly in the five newspapers. Also, a considerable degree of variability between the newspapers in their estimates for the PP and PSOE is shown. It can also be seen that the *ABC* and *El Mundo* newspapers tended to overestimate the PP and underestimate the PSOE, while *El Periódico* and *El País* did just the opposite.

Graphs 5 and 6 are similar, however in the latter, the points corresponding to each party are on the same vertical axis and those of each newspaper are joined by a line. It can be seen that for the CiU, IU and PNV, the five newspapers have a similar trajectory while for the PP and PSOE the lines are clearly different: on the one hand –almost overlapping– those corresponding to *El Mundo* and *ABC* (the newspapers that tend to overestimate the results of the PP and underestimate those of the PSOE); while on the other hand –and also having very closely matched lines– are *El País* and *El Periódico* (the newspapers that tend to overestimate the PSOE

**Graph 4:** Errors in seat prediction for each party, grouped by newspapers and elections



**Graph 5:** Average errors in prediction of the number of seats obtained by each party separating the five evaluated newspapers



results and underestimate those of the PP). *La Vanguardia* has a unique trajectory.

Although the newspaper-party proximity may be argued, in general there is a consensus that considers that the *ABC* and *El Mundo* are more closely tied to the PP<sup>6</sup> while *El País* and *El Periódico* are more linked to the PSOE. *La Vanguardia* cannot be classified in this manner. Therefore, the graphical analysis follows in line with the idea that newspapers tend to be more optimistic regarding the parties with which they have closer ties. However, variance analysis of errors does not reveal the newspaper-party interaction to be

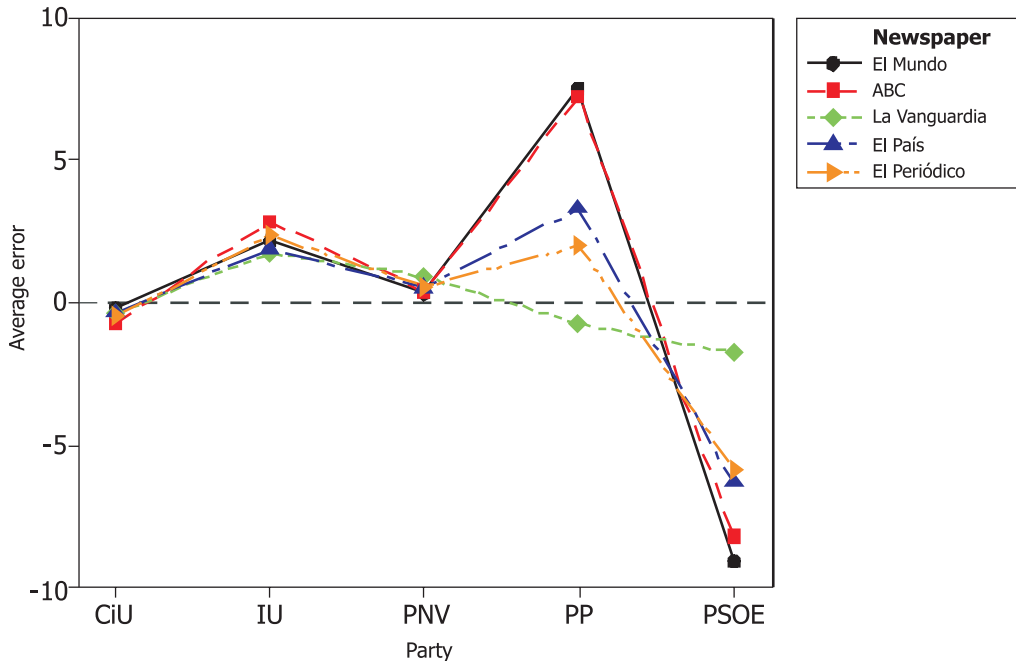
statistically significant, that is, the error committed in predicting the number of seats obtained by a party does not significantly depend on the newspaper publishing it. However, it is true that if this analysis is conducted without including the results of the last elections in which the *ABC* greatly overestimated the PSOE and *El País* the PP (contrary to the other elections), the difference would in fact be clearly significant. We are therefore, faced with a situation in which we cannot make clear affirmations. Annex 3 contains the detailed statistical analyses.

### SOME CONSIDERATIONS REGARDING THE SURVEY DOCUMENTATION ACCOMPANYING THE POLL PUBLICATION

Based on the publication of Law 14/1980, from the 18<sup>th</sup> of April, regarding the regime of elec-

<sup>6</sup> Castromil (2012) conducted a detailed study regarding publication, in the *ABC*, *El Mundo* and *El País* newspapers, of information that could damage a determined party during the election campaign and concluded that the newspapers publishing the most prejudicial information to the PSOE were the *ABC* and *El Mundo* (there is no reliable statistic for other parties).

**Graph 6:** Average errors (in the 8 considered elections) in prediction of the number of seats based on the media in which they are published



toral surveys<sup>7</sup>, all polls published must include “survey documentation” informing, in more or less detail, in regards to the procedures followed to obtain the sample, its size, and the margin of error for a determined confidence level (based on certain assumptions, as usual).

However, the interest value of the provided information is questionable. For example, when informing of a larger sample size (increased effort, higher budget) an increase in prediction quality is expected, however data analysis does not support this expectation. Graph 7 shows the distance between the poll and the actual results in function of the sample size used.

A negative correlation would be expected, that is, an increased sample size would be expected to have a shorter distance to the

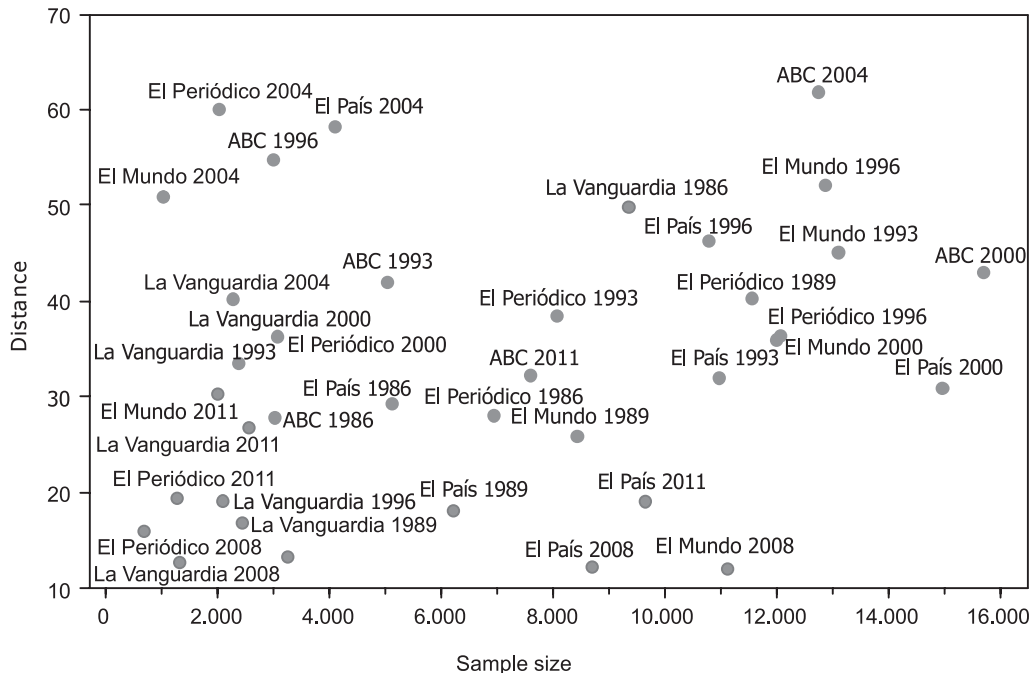
actual result, but this correlation does not exist: the correlation coefficient has a value of 0.2 (positive) with a p-value of 0.272.

There is in fact a relationship between sample size and the margin of error indicated in the survey documentation since one is calculated from the other. Graph 8 shows the margin of error in function of sample size for each of the examined polls. It also shows the mathematical relationship between both variables in the case of a simple random sample, with an infinite population and  $p=q=0.5$ . The similarity between the tendency marked by the points representing the polls and the line corresponding to the simple random sample is quite clear, suggesting that in many cases this formula has been used (only valid for the case of the simple random sample) independent of the sampling strategy that has been used.

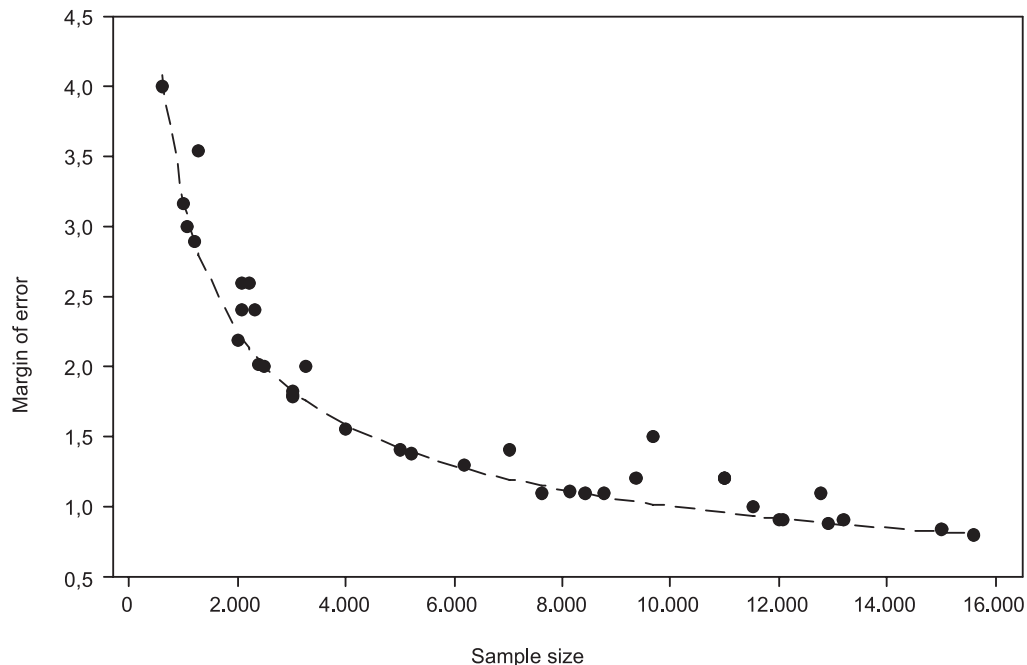
Nor is there a relationship found between prediction quality and margin of error as indi-

<sup>7</sup> A detailed analysis of the background and content of this law may be found in Gálvez (2000).

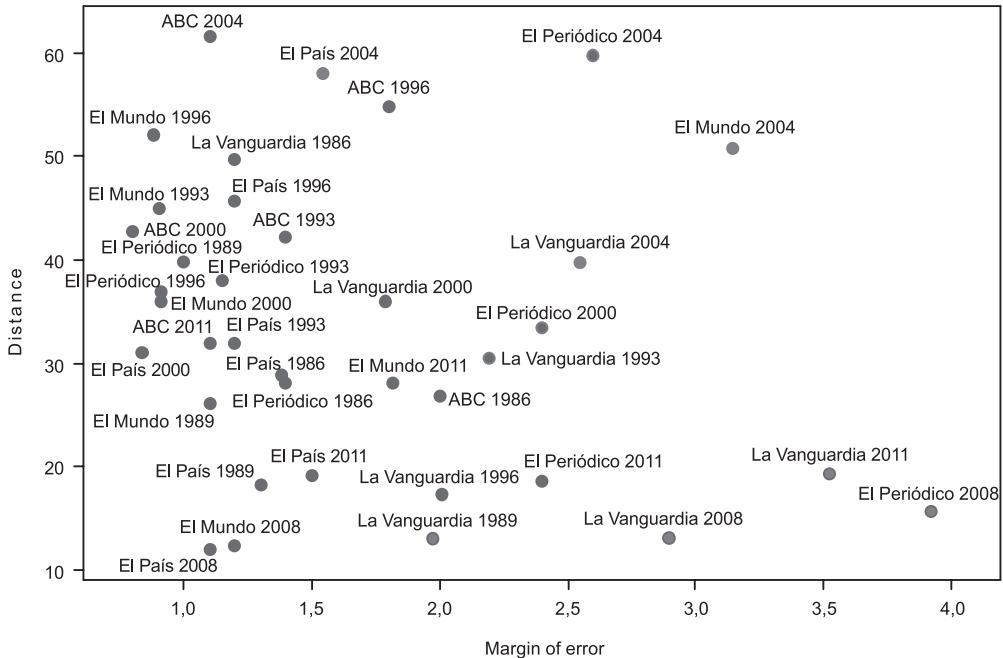
**Graph 7:** Relationship between the distance to the actual result and the sample size of the examined polls



**Graph 8:** Relationship between margin of error and sample size as found in the survey documentation for each poll. The dashed line corresponds to the relationship for a simple random sample, with  $p=q=0.5$



**Graph 9:** Relationship between the distance between the actual result and the margin of error displayed in the survey documentation of the considered polls.



cated in the survey documentation (graph 9). The correlation coefficient between both variables is  $-0.236$  (here negative, though expected to be positive) with a  $p$ -value of  $0.160$ .

Therefore, the data demonstrates that the survey documentation (sample size, margin of error, confidence level, etc.) accompanying the polls, is not useful for evaluating prediction quality, as was intended by the law which requires its publication. Surely, the most relevant information is the name of the company conducting the poll and their accuracy level in previous polls. This level of precision may be measured through a single number, easy to interpret and simple to calculate, such as the poll-actual results distance as used in this study.

## CONCLUSIONS

Despite the difficulties inherent in conducting electoral polls—difficulties extending beyond

those that may be resolved by use of statistical theory—the polls published in Spain one week prior to the general elections have, except in a few cases, been found to have a reasonably high accuracy level.

This conclusion was reached after defining a distance between the predictions made and the actual election results, measuring this distance for the published polls and comparing it with that obtained in the “perfect poll”—a poll that is simulated using a large sample size (larger than any of those of the published polls) and using actual results to generate the selected options of the individuals in the sample. A clear difference is found in the distances when comparing the different elections. The polls offering the worst results were those from the 2004 elections (following the 11M attacks) and the best were those from 2008, in which all of the published polls had distances within the interval expected from the perfect poll. Clear differences are also found between

parties: the party having the best estimates is the CiU and the most difficult parties to estimate are those with the greatest number of representatives (the PSOE and PP), as is expected.

Graphic analysis of the data confirms the idea that newspapers are more optimistic in their predictions regarding results of parties with which they have ties; however, the newspaper-party interaction is not statistically significant based on a more detailed study using variance analysis.

Finally, it has been determined that the survey documentation content serves purely as a formality, offering no information needed to evaluate the quality or reliability of the related polls. The sample selection method and the data "cooking" (using the expression in its best possible sense: assigning votes to the undecided, etc.) forms a part of the company's *know how* in regards to polling. Herein lies the secret to good prediction, and this information is not revealed in any survey documentation.

## REFERENCES

Bouza, Fermín (1998). "Comunicación política: encuestas, agendas y procesos cognitivos electorales". *Praxis Sociológica*, 3: 49-58.

Castromil, Antón R. (2012). "Negativismo mediático y campaña electoral en las Elecciones Generales

de 2008". *Revista Española de Investigaciones Sociológicas*, 139: 163-174.

- Delicado, Pedro and Frederic Udina (2001). "¿Cómo y cuánto fallan los sondeos electorales?". *Revista Española de Investigaciones Sociológicas*, 96: 123-150.
- Díaz de Rada, Vidal (2007). "El trabajo de campo con encuesta presencial. Algunas reflexiones derivadas de la práctica profesional". *Investigación y Marketing*, 96: 11-26.
- Gálvez, Luis (2000). "Organismos de sondeos, encuestas electorales y derecho". *Revista de Estudios Políticos*, 110: 97-121.
- Langsrud, Øyvind (2003). "ANOVA for unbalanced data: Use Type II instead of Type III sums of squares". *Statistics and Computing*, 13: 163-167.
- Manly, Bryan F. J. (1986). *Multivariate Statistical Methods. A primer*. Boca Raton: Chapman & Hall/CRC.
- Pavía, José M. et al (2011). "Encuestas electorales online: nuevos retos, viejos problemas". *Revista Española de Investigaciones Sociológicas*, 135: 107-122.
- Pavía, José M. and Beatriz Larraz (2012). "Sesgo de no-respuesta y modelos de superpoblación en encuestas electorales". *Revista Española de Investigaciones Sociológicas*, 137: 121-150.
- Udina, Frederic and Pedro Delicado (2005). "Estimating Parliamentary composition through electoral polls". *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168: 387-399.

**RECEPTION:** 07/02/2012

**ACCEPTANCE:** 08/02/2013



**Annex 1.** Polls published in the evaluated newspapers and actual results of Congress of Deputies elections from 1986 to 2011.

The last column indicates the date and newspaper page on which the poll was published.

Elections from the 22nd of June of 1986

Polls / Results	PSOE	AP-PDP-PL	CDS	CiU	IU	EAJ-PNV	HB	EE	Others	Date (Page)
ABC	167-180	105-121	13-19	17-20	6-12	8-10	3	1	3	16/06 (20)
<i>El Mundo</i> (1)	-	-	-	-	-	-	-	-	-	-
<i>El País</i>	167-194	86-105	15-29	17-20	10-16	8-9	3-4	1-2	4-7	15/06 (11)
<i>El Periódico</i>	180	98	30	17	5	7	4	1	8	15/06 (3)
<i>La Vanguardia</i>	178-199	71-90	17-28	20-21	7-12	9-11	1-2	1-2	11-17	16/06 (5)
<b>Actual results</b>	<b>184</b>	<b>105</b>	<b>19</b>	<b>18</b>	<b>7</b>	<b>6</b>	<b>5</b>	<b>2</b>	<b>4</b>	

Elections from the 29th of October of 1989

Polls / Results	PSOE	PP	CiU	IU	CDS	EAJ-PNV	HB	PA	Others	Date (Page)
ABC (2)	-	-	-	-	-	-	-	-	-	-
<i>El Mundo</i>	168	103	17	16	23	5	5	2	11	23/10 (1)
<i>El País</i>	176-182	95-104	18	15-16	16-19	5-7	4-5	2-3	6-8	22/10 (13)
<i>El Periódico</i>	172-178	87-91	17-19	22-26	18-21	7	7	4	7-10	22/10 (3)
<i>La Vanguardia</i>	172-178	104-107	17-18	13	14-15	7-8	3	3-4	5-12	23/10 (1)
<b>Actual results</b>	<b>175</b>	<b>107</b>	<b>18</b>	<b>17</b>	<b>14</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>8</b>	

Elections from the 6th of June of 1993

Polls / Results	PSOE	PP	IU	CiU	EAJ-PNV	CC	HB	ERC	Others	Date (Page)
ABC	134-144	144-153	26-28	16-17	6-7	3-4	3	1	5-7	29/05 (23)
<i>El Mundo</i>	132-141	147-155	25-27	17-19	5-6	3	3-4	1	5-6	30/05 (7)
<i>El País</i>	135-151	141-155	20-22	18-20	6	3	2-3	1	5-8	30/05 (13)
<i>El Periódico</i>	133-152	134-151	19-24	19-21	6-7	4-6	3-4	1	6-13	30/05 (4)
<i>La Vanguardia</i>	142-146	144-148	24-25	18-19	6-7	*	*	0-1	9-11	30/05 (18)
<b>Actual results</b>	<b>159</b>	<b>141</b>	<b>18</b>	<b>17</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>1</b>	<b>3</b>	

Elections from the 3rd of March of 1996

Polls / Results	PP	PSOE	IU	CiU	EAJ-PNV	CC	BNG	HB	Others	Date (Page)
ABC	176-184	117-125	22-28	13-14	4-5	*	*	*	6-8	25/02 (5)
<i>El Mundo</i>	170-179	113-123	25-29	14-15	6-7	2-3	2	2	3	25/02 (9)
<i>El País</i>	170-178	118-128	24-27	13-15	5-6	3	1	2	2-3	25/02 (16)
<i>El Periódico</i>	165-175	119-129	18-27	14-16	5	2-3	1	2	2-5	20/02 (5)
<i>La Vanguardia</i>	160-170	135-145	19-21	14-15	5-6	*	*	*	7-8	25/02 (13)
<b>Actual results</b>	<b>156</b>	<b>141</b>	<b>21</b>	<b>16</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>2</b>	<b>3</b>	

(1) As of this date, the *El Mundo* newspaper was not yet in existence.

(2) For the 1989 elections the newspaper, ABC did not publish any polls.

\* Included in *Others*.

**Annex 1.** Polls published in the evaluated newspapers and actual results of Congress of Deputies elections from 1986 to 2011.(continuación)

Elections from the 12th of March of 2000

Polls / Results	PP	PSOE	CiU	IU	EAJ-PNV	CC	BNG	PA	Others	Date (Page)
<i>ABC</i>	158-164	138-144	15-16	9-12	6-7	4	3-5	*	3-4	05/03 (1)
<i>El Mundo</i>	164-170	137-143	16	9-11	6-7	4	3-5	0-1	2	05/03 (9)
<i>El País</i>	165-171	131-139	14-15	9-13	6-7	4	4-5	0	3-5	05/03 (17)
<i>El Periódico</i>	165-170	135-140	13-14	9-12	7-8	4-5	4-5	2	1	05/03 (4)
<i>La Vanguardia</i>	166-170	136-140	14-15	12-14	6	4	3	*	2-5	05/03 (1)
<b>Actual results</b>	<b>183</b>	<b>125</b>	<b>15</b>	<b>8</b>	<b>7</b>	<b>4</b>	<b>3</b>	<b>1</b>	<b>4</b>	

Elections from the 14th of March of 2004

Polls / Results	PSOE	PP	CiU	ERC	EAJ-PNV	IU	CC	BNG	Others	Date (Page)
<i>ABC</i>	133-137	174-177	10-11	6-7	7	8-10	3	2	2	29/02 (1)
<i>El Mundo</i>	138-144	168-173	11	5-6	7	6-8	3-4	3	1-2	08/03 (8)
<i>El País</i>	134-141	168-172	10-11	7	7-8	11	3-4	3	2	07/03 (14)
<i>El Periódico</i>	135-140	169-173	10-11	6-7	7	10-12	*	*	10-14	07/03 (3)
<i>La Vanguardia</i>	143-147	162-167	10	5-6	7	9-10	*	*	4-5	07/03 (1)
<b>Actual results</b>	<b>164</b>	<b>148</b>	<b>10</b>	<b>8</b>	<b>7</b>	<b>5</b>	<b>3</b>	<b>2</b>	<b>3</b>	

Elections from the 9th of March of 2008

Polls / Results	PSOE	PP	CiU	EAJ-PNV	ERC	IU	BNG	CC	Others	Date (Page)
<i>ABC</i> (3)	-	-	-	-	-	-	-	-	-	02/03 (14)
<i>El Mundo</i>	157-171	148-161	9-11	7	5-6	4	0	2	2	03/03 (10)
<i>El País</i>	165-169	148-154	9	7	5-6	4	2	2	2-3	02/03 (14)
<i>El Periódico</i>	165-170	149-154	8-9	6	5-6	4-5	*	*	8-11	28/02 (18)
<i>La Vanguardia</i>	162-167	152-156	8-9	6-7	6	4-5	*	*	5-7	03/03 (14)
<b>Actual results</b>	<b>169</b>	<b>154</b>	<b>10</b>	<b>6</b>	<b>3</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	

Elections from the 20th of November of 2011

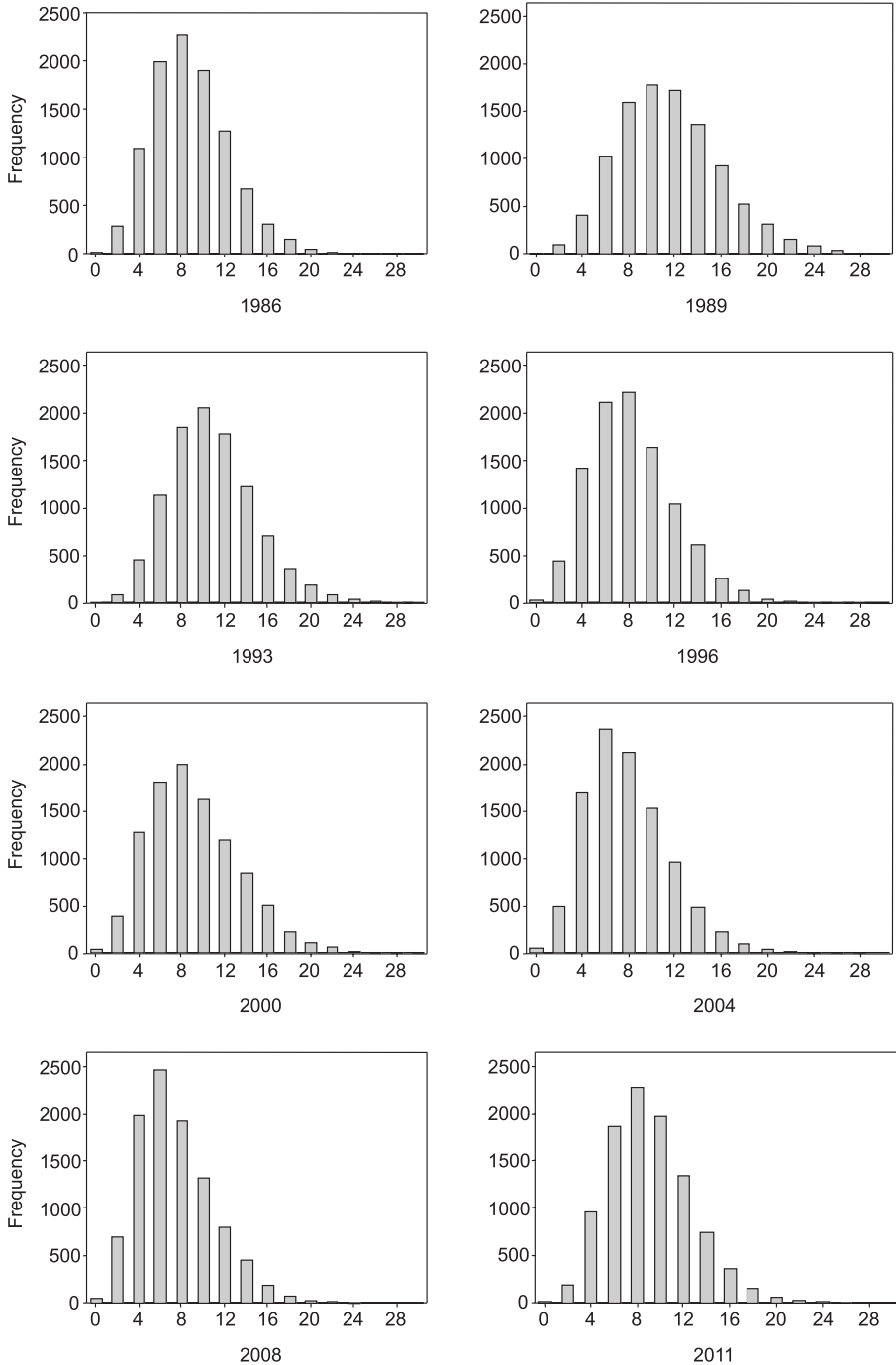
Polls / Results	PP	PSOE	CiU	IU	AMAIUR	UPyD	EAJ-PNV	ERC	Others	Date (Page)
<i>ABC</i>	187-188	123-126	12-14	6-7	5-6	2-3	4	2	5-6	13/11 (19)
<i>El Mundo</i>	198	112	14	7	3	3	5	2	6	13/11 (4)
<i>El País</i>	192-196	110-113	14	11	4-5	2	5	3	5	13/11 (14)
<i>El Periódico</i>	188-140	115-118	13-14	8-10	4-5	4	4-5	2-3	7-8	13/11 (31)
<i>La Vanguardia</i>	184-147	116-120	12-14	8-10	3-4	3-4	4-5	3	8-9	13/11 (17)
<b>Actual results</b>	<b>186</b>	<b>110</b>	<b>16</b>	<b>11</b>	<b>7</b>	<b>5</b>	<b>5</b>	<b>3</b>	<b>7</b>	

(3) Predictions are offered in percentages but not in number of seats.

\* Included in *Others*.

**Annex 2.** *Distribution of distance values in the perfect poll simulations*

**Graph A2.1:** *Distribution of distances obtained upon conducting 10,000 repetitions of the perfect poll for each of the evaluated elections*



**Annex 3.** Table of errors for each party (difference between the mean value of the prediction interval and the actual value) for the evaluated polls

		PSOE	PP*	CiU	IU	PNV
1986	<i>ABC</i>	-10.5	5	0.5	2	3
	<i>El Mundo</i>	—	—	—	—	—
	<i>El País</i>	-3.5	-6.5	0.5	-2	1
	<i>El Periódico</i>	-4	4	-1	-2	1
	<i>La Vanguardia</i>	4.5	-21	2.5	2.5	4
1989	<i>ABC</i>	-	-	-	-	-
	<i>El Mundo</i>	-7	5	-1	-1	0
	<i>El País</i>	4	-4	0	-1.5	1
	<i>El Periódico</i>	0	-12.5	0	7	2
	<i>La Vanguardia</i>	0	-1	-0.5	-4	2.5
1993	<i>ABC</i>	-20	7.5	-0.5	9	1.5
	<i>El Mundo</i>	-22.5	10	1	8	0.5
	<i>El País</i>	-16	7	2	3	1
	<i>El Periódico</i>	-16.5	1.5	3	3.5	1.5
	<i>La Vanguardia</i>	-15	5	1.5	6.5	1.5
1996	<i>ABC</i>	-20	24	-2.5	4	-0.5
	<i>El Mundo</i>	-23	18.5	-1.5	6	1.5
	<i>El País</i>	-18	18	-2	4.5	0.5
	<i>El Periódico</i>	-17	14	-1	1.5	0
	<i>La Vanguardia</i>	-1	9	-1.5	-1	0.5
2000	<i>ABC</i>	16	-22	0.5	2.5	-0.5
	<i>El Mundo</i>	15	-16	1	2	-0.5
	<i>El País</i>	10	-15	-0.5	3	-0.5
	<i>El Periódico</i>	12.5	-15.5	-1.5	2.5	0.5
	<i>La Vanguardia</i>	13	-15	-0.5	5	-1
2004	<i>ABC</i>	-29	27.5	0.5	4	0
	<i>El Mundo</i>	-23	22.5	1	2	0
	<i>El País</i>	-26.5	22	0.5	6	0.5
	<i>El Periódico</i>	-26.5	23	0.5	6	0
	<i>La Vanguardia</i>	-19	16.5	0	4.5	0
2008	<i>ABC</i>	-	-	-	-	-
	<i>El Mundo</i>	-5	0.5	0	2	1
	<i>El País</i>	-2	-3	-1	2	1
	<i>El Periódico</i>	-1.5	-2.5	-1.5	2.5	0
	<i>La Vanguardia</i>	-4.5	0	-1.5	2.5	0.5
2011	<i>ABC</i>	14.5	1.5	-3	-4.5	-1
	<i>El Mundo</i>	2	12	-2	-4	0
	<i>El País</i>	1.5	8	-2	0	0
	<i>El Periódico</i>	6.5	4	-2.5	-2	-0.5
	<i>La Vanguardia</i>	8	0.5	-3	-2	-0.5

\* In the 1986 elections, the seats assigned to the PP were those of the AP-PDP-PL coalition together with those obtained by the CDS. In the 1989 elections, they were those of both the PP and those of the CDS.

The existence of missing values (*El Mundo* in 1986 and *ABC* in 1989 and 2008) hinders the data analysis process for determining the statistical significance of the newspaper-party interaction through variance analysis. A reasonable manner of resolving this difficulty is to assign to each missing value the average of the values existing for this same newspaper and party. In this way, for example, the missing value corresponding to the error committed by *El Mundo* in the prediction of the number of seats obtained by the PSOE in the 1986 elections has been substituted by the average of the other errors also committed by *El Mundo* with the PSOE, that is: -7 (in 1989), -22.5 (1993), -23 (1996), 15 (2000), -23 (2004), -5 (2008) and 2 (2011), resulting in a value of -9.07. The other missing values have been treated in this same manner<sup>1</sup>.

Having solved the problem of the missing values, the study now focused on those errors committed regarding the PSOE and PP by newspapers that are considered to have ties to one party or the other. In the variance analysis, a fixed effects model was used having the following three factors: Year, Newspaper and Party and the following three interactions: Year-Newspaper, Year-Party and Newspaper-Party, with this last interaction being the one of greatest interest. The following results were obtained using the MINITAB statistics software package:

General Linear Model: Error versus Year; Newspaper; Party						
Factor	Type	Levels	Values			
Year	fixed	7	1986; 1989; 1993; 1996; 2000; 2004; 2008			
Newspaper	fixed	4	ABC; El Mundo; El País; El Periódico			
Party	fixed	2	PP; PSOE			
Analysis of Variance for Error, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Year	7	660.41	660.41	94.34	3.13	0.020
Newspaper	3	20.87	20.87	6.96	0.23	0.874
Party	1	2442.95	2442.95	2442.95	81.14	0.000
Year*Newspaper	21	121.29	121.29	5.78	0.19	1.000
Year*Party	7	8822.33	8822.33	1260.33	41.86	0.000
Newspaper*Party	3	220.97	220.97	73.66	2.45	0.092 ←
Error	21	632.30	632.30	30.11		
Total	63	12921.12				
S = 5.48722 R-Sq = 95.11% R-Sq(adj) = 85.32%						

<sup>1</sup> It may also be resolved by calculating the variance analysis table using Type II sums of squares, as recommended by Langsrud (2003). We used the R statistical software with the ANOVA function of the "car" package and the results are practically the same.

The Newspaper·Party interaction is found to have a p-value of de 0.092 and, therefore, with the usual criteria of a 5% significance level, it is not considered to be statistically significant. However, when excluding the results from the last elections held in November of 2011, the interaction is in fact found to be significant:

General Linear Model: Error (sin 2011) versus Year; Newspaper; Party

Factor	Type	Levels	Values
Year	fixed	7	1986; 1989; 1993; 1996; 2000; 2004; 2008;
Newspaper	fixed	4	ABC; El Mundo; El País; El Periódico
Party	fixed	2	PP; PSOE

Analysis of Variance for Error, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Year	6	158.03	158.03	26.34	1.14	0.378
Newspaper	3	13.18	13.18	4.39	0.19	0.901
Party	1	2777.84	2777.84	2777.84	120.53	0.000
Year*Newspaper	18	115.23	115.23	6.40	0.28	0.995
Year*Party	6	8487.32	8487.32	1414.55	61.38	0.000
Newspaper*Party	3	279.82	279.82	93.27	4.05	0.023 ←
Error	18	414.83	414.83	23.05		
Total	55	12246.25				

S = 4.80064    R-Sq = 96.61%    R-Sq(adj) = 89.65%

**Annex 4.** Information obtained from the survey documentation of the analyzed polls

	Company producing the poll	Sample size	Margin of error	Date and page of publication	
1986	<i>ABC</i>	Aresco, SA	2,505	2.00	16/06 p. 21
	<i>El Mundo</i>	–	–	–	–
	<i>El País</i>	Inst. Demoscopia	5,200	1.38	15/06 p. 16
	<i>El Periódico</i>	IOPE-ETMAR	7,029	1.40	15/06 p. 4
	<i>La Vanguardia</i>	Line Staff	9,371	1.20	16/06 p. 15
1989	<i>ABC</i>	–	–	–	–
	<i>El Mundo</i>	Sigma Dos	8,400	1.10	23/10 p. E12
	<i>El País</i>	Inst. Demoscopia	6,200	1.30	22/10 p. 18
	<i>El Periódico</i>	ICP/Research	11,524	1.00	22/10 p. 4
	<i>La Vanguardia</i>	Inst. Opina	3,262	2.00	23/10 p. 13
1993	<i>ABC</i>	Gruppo, SA	5,000	1.40	29/05 p. 23
	<i>El Mundo</i>	Sigma Dos	13,200	0.90	30/05 p. 13
	<i>El País</i>	Demoscopia SA	11,000	1.20	30/05 p. 14
	<i>El Periódico</i>	Vox Publica / IOPE Etmar	8,148	1.11	30/05 p. 4
	<i>La Vanguardia</i>	Inst. Opina	2,000	2.19	30/05 p. 21
1996	<i>ABC</i>	Tabula V	3,000	1.80	25/02 p. 29
	<i>El Mundo</i>	Sigma Dos & Vox Publica	12,900	0.88	25/02 p. 15
	<i>El País</i>	Demoscopia	11,000	1.20	25/02 p. 17
	<i>El Periódico</i>	Vox Publica	12,069	0.91	20/02 p. 6
	<i>La Vanguardia</i>	Inst. Opina	2,369	2.01	25/02 p. 17
2000	<i>ABC</i>	Ipsos - Ecoconsulting	15,600	0.80	05/03 p. 21
	<i>El Mundo</i>	Sigma Dos	12,000	0.91	05/03 p. 9
	<i>El País</i>	Demoscopia	15,000	0.84	05/03 p. 19
	<i>El Periódico</i>	Vox Publica	2,300	2.40	05/03 p. 6
	<i>La Vanguardia</i>	Inst. Opina	3,000	1.79	05/03 p. 20
2004	<i>ABC</i>	TSN-Demoscopia	12,760	1.10	29/02 p. 15
	<i>El Mundo</i>	Sigma Dos	1,000	3.16	08/03 p. 8
	<i>El País</i>	Inst. Opina	4,000	1.55	07/03 p. 15
	<i>El Periódico</i>	Vox Publica	2,071	2.60	07/03 p. 9
	<i>La Vanguardia</i>	Inst. Noxa	2,200	2.60	07/03 p. 16
2008	<i>ABC</i>	Inst. DYM	1,075	3.00	02/03 p. 14
	<i>El Mundo</i>	Sigma Dos	11,000	1.20	03/03 p. 11
	<i>El País</i>	Metroscopia	8,750	1.10	02/03 p. 14
	<i>El Periódico</i>	GESOP	600	4.00	28/02 p. 18
	<i>La Vanguardia</i>	Inst. Noxa	1,200	2.89	03/03 p. 16
2011	<i>ABC</i>	Inst. DYM	7,612	1.10	02/03 p. 14
	<i>El Mundo</i>	Sigma Dos	3,000	1.82	03/03 p. 11
	<i>El País</i>	Metroscopia	9,675	1.50	02/03 p. 14
	<i>El Periódico</i>	GESOP	2,070	2.40	28/02 p. 18
	<i>La Vanguardia</i>	Inst. Noxa	1,275	3.54	03/03 p. 16



# ¿Aciertan los sondeos electorales? Análisis sobre la bondad de predicción de los sondeos electorales publicados en la prensa

*Are Election Polls Right? Analysis of the Accuracy of Election Polls Predictions Published in the Press*

**Adrià Caballé, Pere Grima y Lluís Marco-Almagro**

## Palabras clave

Encuestas • Elecciones  
• Trabajo de campo  
• Conducta electoral  
• Medios de comunicación  
• Distribución de escaños • Diseño de la investigación

## Key words

Surveys • Elections  
• Fieldwork • Electoral Behaviour • Mass Media • Distribution of Parliamentary Seats  
• Research Design

## Resumen

Este artículo analiza las predicciones sobre los resultados de las elecciones al Congreso de los Diputados de España publicadas en los principales medios de la prensa española. Para comparar su grado de acierto se define una distancia entre la previsión y los resultados reales, y se valora su magnitud comparándola con la que presenta lo que se ha denominado “sondeo perfecto”, realizado por simulación a partir de las proporciones realmente obtenidas por cada partido. También se analizan otros aspectos relacionados con los sondeos publicados, como la relación entre el tamaño de muestra y la calidad de la estimación realizada, o entre el error que se comete (por exceso o por defecto) con un partido y la afinidad ideológica entre dicho partido y el medio que publica la encuesta.

## Abstract

This paper analyses the predictions of the election results for the Spanish Parliament published in the main Spanish newspapers. To gauge their degree of accuracy, the distance between the predictions and the actual results was defined and later assessed by comparing it with that of the so-called ‘perfect poll’. This ‘perfect poll’ was generated by a simulation using the proportions of votes actually obtained by each party. Other issues related to the published polls are also analysed, such as the relationship between sample size and estimation quality, and between the error made (by excess or defect) for a party and the ideological affinity between that party and the newspaper publishing the survey.

## Cómo citar

Caballé, Adrià; Pere Grima y Lluís Marco-Almagro (2013). «¿Aciertan los sondeos electorales? Análisis sobre la bondad de predicción de los sondeos electorales publicados en la prensa». *Revista Española de Investigaciones Sociológicas*, 143: 25-46. (<http://dx.doi.org/10.5477/cis/reis.143.25>)

La versión en inglés de este artículo puede consultarse en <http://reis.cis.es> y <http://reis.metapress.com>

**Adrià Caballé:** Universidad Politécnica de Cataluña | [adriacaballe@yahoo.es](mailto:adriacaballe@yahoo.es)

**Pere Grima:** Universidad Politécnica de Cataluña | [pere.grima@upc.edu](mailto:pere.grima@upc.edu)

**Lluís Marco-Almagro:** Universidad Politécnica de Cataluña | [lluis.marco@upc.edu](mailto:lluis.marco@upc.edu)

## INTRODUCCIÓN<sup>1</sup>

Siempre ha existido cierta controversia sobre la capacidad de predicción de los sondeos electorales. Por una parte, es habitual referirse a ellos en un tono de incredulidad («solo son encuestas»), especialmente por aquellos a los que no les son favorables, pero, por otra, se han convertido en un elemento habitual en la prensa y son utilizados como guía para los comentarios y para la interpretación de la realidad política del momento.

Se pueden considerar dos tipos de estudios sobre intención de voto: los que se realizan a lo largo de la legislatura presentándose bajo titulares del tipo «si las elecciones se celebraran ahora...», y los que se publican una semana antes de las elecciones. En el primer caso es imposible saber si los resultados que se dan tienen o no un nivel de aproximación razonable, siendo esto lo que ocurre en general con los estudios basados en el análisis de muestras (como los estudios de las audiencias de TV, por ejemplo). Sin embargo, para los que se publican una semana antes de las elecciones sí que se acaba sabiendo «la verdad», y por tanto se puede valorar de forma objetiva la predicción realizada.

Además de esta peculiaridad, y de que muchas personas identifican las posibilidades de la estadística con lo acertadas que resultan este tipo de predicciones, lo cierto

es que los sondeos electorales son un ejercicio lleno de dificultades, algunas de las cuales van más allá de lo que son capaces de abordar las técnicas estadísticas, y que se pueden resumir en los siguientes puntos:

— *Dificultad de obtener una muestra representativa.* Esto no es algo específico de los sondeos electorales, cualquier predicción basada en el estudio de una muestra exige que esta sea representativa y conseguirlo no es fácil ni barato. La ficha técnica de muchos sondeos electorales indica que las encuestas se han realizado por teléfono, y en este caso tiene mucha importancia a qué hora se llama, por quién se pregunta, o cómo se sustituye a quienes no desean contestar. Los conocimientos y experiencia de las empresas especializadas para conseguir una muestra representativa a precio competitivo es una de las claves del éxito. Un listado de dificultades inherentes a toda encuesta presencial puede encontrarse en Díaz de Prada (2007). Las que se realizan online tienen también sus propios problemas (Pavía et al., 2011).

— *La intención de voto va cambiando.* Los sondeos electorales que se publican una semana antes de las elecciones se basan en encuestas realizadas varios días, o incluso alguna semana, antes de su publicación. Por tanto, además de la extrapolación que se hace de la muestra a la población (siendo esta la que trata la teoría estadística del muestreo), también se generaliza, de las fechas en que se ha realizado la encuesta, al día de las elecciones, y en ese periodo pueden ocurrir sucesos que afecten al voto decidido o a la decisión final de los que estaban indecisos.

— *Asignación de voto a los indecisos.* El problema está en que las personas que no responden no son una muestra al azar de la población. Si las razones de no respuesta fueran estadísticamente independientes de las preferencias de voto el problema tendría

<sup>1</sup> Nuestro interés por este tema surgió a partir de nuestras discusiones con los profesores Roberto Behar, de la Universidad del Valle en Cali (Colombia), y Pedro Delicado, de la Universitat Politècnica de Catalunya (UPC), a los cuales queremos manifestar nuestro agradecimiento. También agradecemos al profesor Josep-Antón Sánchez-Espigares, de la UPC, su ayuda en algunos aspectos estadísticos y computacionales del estudio. A Víctor Peña, estudiante del máster en Estadística e Investigación Operativa de la UPC, por su detallada lectura del original gracias a la cual pudimos corregir algún error. Finalmente, a dos evaluadores anónimos de la revista que realizaron un detallado análisis de nuestro trabajo y con sus sugerencias nos permitieron mejorarlo notablemente.

fácil solución desde la estadística. Sin embargo, la práctica ha demostrado que esto no es así y el problema se complica, tal como explica Bouza (1998).

— *Falta de respuesta o de sinceridad.* A algunas personas no les gusta manifestar sus preferencias políticas (podría decirse que más que indecisos son «decisos cautos»). Además, esta reserva a manifestar la intención de voto se da más entre los que piensan votar a determinados partidos. Pavía y Larráz (2012) han analizado a fondo este problema que definen como «la principal fuente de error de las predicciones electorales en España» y plantean estrategias para abordarlo.

— *Asignación del número de escaños en base al porcentaje de votos.* Las leyes estadísticas que se manejan sirven para estimar la proporción de votos, pero lo verdaderamente relevante es el número de escaños. Si en una determinada circunscripción hay 5 escaños en juego y se puede predecir con una confianza del 95% que un determinado partido obtendrá el 32% de los votos con un margen de error del 3%, el problema está en que si obtiene un 31% le corresponderá un escaño, mientras que si obtiene el 33% le corresponderán dos. Esta es una diferencia importante, pero la información disponible no permite decantarse por una opción u otra. Otra dificultad la introduce la necesidad de alcanzar el 3% de los votos para entrar en el reparto de escaños; algunos partidos pueden rondar esa proporción, siendo imposible saber si la alcanzarán o no, cosa que puede repercutir en el reparto general. Finalmente, la aplicación de la fórmula D'Hondt también introduce algunos problemas de estimación: mientras que la estimación de la proporción de votos es insesgada (el promedio de las proporciones obtenidas si se realizaran muchos sondeos coincidiría con las verdaderas), el número de escaños obtenidos aplicando la fórmula D'Hondt da una estimación sesgada, tal como ponen de manifiesto Delicado y Udina (2001).

El objetivo principal de este artículo es responder a la pregunta planteada en el título aportando argumentos basados en el análisis de los datos. Naturalmente, el término «acertar» no hay que tomarlo al pie de la letra (ya sabemos que no se acierta «exactamente»), sino en el sentido de si las predicciones realizadas pueden considerarse razonablemente aproximadas teniendo en cuenta el margen de error que matemáticamente conlleva cualquier estimación estadística<sup>2</sup>.

## SONDEOS ANALIZADOS

Se han analizado los sondeos para las elecciones al Congreso de los Diputados publicados en los diarios de mayor difusión en España: *El País*, *El Mundo*, *ABC*, *La Vanguardia* y *El Periódico*<sup>3</sup>, abarcando el periodo que va desde las elecciones del 22 de junio de 1986 hasta las del 20 de noviembre de 2011, ya que para las tres primeras elecciones desde el retorno de la democracia (15 de junio de 1977, 1 de marzo de 1979 y 28 de octubre de 1982) no se publicaron sondeos en los medios considerados.

El Anexo 1 contiene los resultados de los sondeos analizados, ordenando los partidos según el número de escaños que obtuvieron, así como la fecha y la página del diario en que fueron publicados.

<sup>2</sup> En aras del rigor terminológico debemos aclarar que una cosa son los resultados que se obtienen aplicando métodos estadísticos a los datos obtenidos en el muestreo (y ese sería *stricto sensu* el resultado del sondeo) y otra son los resultados que se publican después de que se hayan corregido de acuerdo con los criterios de la empresa que lo realiza. El análisis que aquí se presenta se refiere a los resultados publicados, que siempre incluyen los ajustes que las empresas responsables han incorporado. Por otra parte, este es el único análisis posible, puesto que las respuestas directas de los ciudadanos incluidos en la muestra nunca se publican.

<sup>3</sup> Oficina de Justificación de Medios (OJD): <http://www.ojd.es/OJD/> (consultado el 24 de noviembre de 2011).

## MEDIDA DE LA CALIDAD DE LA PREDICCIÓN

Para valorar y también para poder realizar comparaciones entre distintos sondeos es necesario definir una medida que cuantifique la calidad de la predicción. Entre las muchas posibilidades que existen (puede consultarse, por ejemplo, Manly, 1986) se pueden considerar las siguientes.

### Distancia agregada (DA)

Es igual a la suma de las diferencias, en valor absoluto, entre las predicciones realizadas y los valores obtenidos por cada partido. Es decir, sean  $e_1, e_2, \dots, e_k$  los escaños obtenidos por los  $k$  partidos que se presentan a una contienda electoral y  $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_k$  las previsiones publicadas en un determinado sondeo. La *distancia agregada* entre las previsiones y la realidad será:

$$DA = \sum_{i=1}^k |e_i - \hat{e}_i|$$

Por ejemplo, las predicciones que publicó *El Periódico* para las elecciones del 22 de junio de 1986 y los valores que realmente se obtuvieron fueron los siguientes:

	PSOE	AP	CDS	CiU	IU	PNV	HB	EE	Otros
<i>El Periódico</i>	180	98	30	17	5	7	4	1	8
Resultados reales	184	105	19	18	7	6	5	2	4
Diferencia (valor absoluto)	4	7	11	1	2	1	1	1	4

Por tanto, la *DA* entre esta predicción y la realidad será:

$$DA = 4 + 7 + 11 + 1 + 2 + 1 + 1 + 1 + 4 = 32$$

### Distancia euclídea (DE)

Los resultados obtenidos por los  $k$  partidos se pueden considerar como un vector en un espacio de  $k$  dimensiones y la *distancia euclídea* es la distancia ordinaria en este tipo de espacios, siendo su expresión una generalización del teorema de Pitágoras:

$$DE = \sqrt{\sum_{i=1}^k (e_i - \hat{e}_i)^2}$$

### Distancia euclídea estandarizada (DEE)

El error —en valor absoluto— en la estimación del número de escaños tiende a ser mayor en los partidos que obtienen más diputados. Así, por ejemplo, es más razonable cometer un error de  $\pm 4$  diputados en la estimación de los resultados del PSOE en las elecciones de 1986, en que obtuvo 184, que en la estimación de IU, que obtuvo 7, o en la de EE, que obtuvo 2 (en este último caso, por defecto, solo se puede cometer un error de 2).

La razón de que se dé esta circunstancia estriba en que la variabilidad del estimador del porcentaje de votos va creciendo a medida que aumenta dicho porcentaje hasta llegar al máximo cuando este es del 50%. Por tanto, es razonable pensar en estandarizar la diferencia entre el valor estimado y el valor real con una

medida de la variabilidad de la estimación que dependerá de la proporción de votos obtenida por cada partido. Considerando, para simplificar, que las variabilidades

son independientes, la expresión de la *distancia euclídea estandarizada* queda:

$$DEE = \sqrt{\sum_{i=1}^k \left( \frac{e_i - \hat{e}_i}{\sigma_i} \right)^2}$$

siendo  $\sigma_i$  la desviación estándar del estimador del número de escaños obtenido por el partido  $i$ . Si un partido obtiene una proporción de votos  $\rho_i$ , la varianza del estimador de dicha proporción responde a la expresión:  $V(\hat{\rho}_i) = \frac{\rho_i(1-\rho_i)}{n}$  siendo  $n$  el tamaño de la muestra. Sin embargo, la varianza del estimador del número de escaños no es tan fácil (la fórmula D'Hondt destruye sus propiedades matemáticas, véase Udina y Delicado, 2005) y habría que estimarlo por simulación. También podrían incluirse las covarianzas entre los distintos partidos obteniendo la llamada Distancia de Mahalanobis, pero a efectos de lo que aquí nos interesa no creemos que ese sea el camino adecuado.

Comparando estas distancias, está claro que la más sencilla y fácil de interpretar es la *distancia agregada*. La distancia euclídea produce una ordenación de las predicciones similar a la distancia agregada (a mayor dife-

los errores en los partidos pequeños, cuando lo que más interesa son los resultados de los partidos grandes, que son los que marcarán el color del nuevo gobierno. Así pues, se ha considerado que la *distancia agregada*, además de sus méritos en cuanto a claridad y facilidad de comprensión, tiene un comportamiento que se adapta bien a nuestros intereses en una situación como la que estamos estudiando, y esta es la que vamos a utilizar.

## DISTANCIA ENTRE SONDEOS Y RESULTADOS

Calcular la distancia presenta, de entrada, dos dificultades que es necesario abordar: el uso de intervalos para dar las predicciones y el hecho de que la categoría «otros» no siempre incluya el mismo número de partidos.

El problema de los intervalos se resuelve generando todas las combinaciones de resultados posibles y calculando la distancia solo para aquellas que suman los 350 diputados que deben salir. El promedio de esas distancias, que corresponden a resultados factibles, es el valor asignado como distancia entre la previsión y el resultado real.

Por ejemplo, el sondeo publicado por *El País* para las últimas elecciones de noviembre de 2011 fue:

	PP	PSOE	CiU	IU-LV	AMAIUR	UPyD	EAJ-PNV	ERC	Otros
Sondeo <i>El País</i>	192-196	110-113	14	11	4-5	2	5	3	5
Resultados reales	186	110	16	11	7	5	5	3	7

rencia en valor absoluto, mayor diferencia al cuadrado), pero en nuestro caso, con  $n$  partidos, no es tan intuitiva. La *distancia euclídea estandarizada* puede considerarse más justa porque valora los errores atendiendo a la probabilidad de que se produzcan, pero tiene el inconveniente de que penaliza más

En este caso, para el PP se presentan 5 posibles resultados, 4 para el PSOE, 2 para AMAIUR y 1 para el resto, de forma que el número total de combinaciones es:  $5 \times 4 \times 2 = 40$ , pero solo se pueden considerar 9, ya que solo esas suman 350. El promedio de las distancias correspondientes a esas 9 com-

binaciones de resultados factibles será la distancia asignada a este sondeo (en este caso es 19).

Obsérvese que la única forma de obtener una distancia igual a cero sería acertando el resultado para todos los partidos con estimaciones concretas, no mediante intervalos, y que el método propuesto penaliza los intervalos anchos, ya que aunque lleguen a incluir el resultado verdadero, también incluirán muchos falsos que tenderán a aumentar la distancia.

Respecto a la segunda dificultad (no todos los sondeos publican sus previsiones para el mismo número de partidos), si la distancia se calcula utilizando siempre todas las estimaciones que da el sondeo se penaliza a los que presentan mayor nivel de detalle (incluyen más partidos) y esto no es razonable. Veamos ahora el sondeo publicado por *El Mundo* para las elecciones de 2008:

tegoría), la previsión coincidiría con la realidad (previsión = realidad = 8) y tendríamos una distancia menor.

Para que no tengan ventaja los sondeos que realizan una mayor agrupación en «Otros» y para que los resultados sean comparables, las distancias se han calculado con las previsiones realizadas para los 5 partidos que obtienen mayor número de escaños, agrupando el resto en la categoría «Otros». De esta forma, además de aplicar un criterio homogéneo, se pueden incluir en el estudio todos los sondeos publicados en el periodo considerado (tabla I).

Estos resultados se han representado en el gráfico 1, donde se observa que las previsiones para las elecciones de 2008 fueron las mejores y con un grado de acierto muy similar en todos los medios considerados, mientras que las de 2004 (posteriores a los atentados del 11-M) fueron las

	PSOE	PP	CiU	EAJ-PNV	ERC	IU	BNG	CC	Otros
Sondeo <i>El Mundo</i>	157-171	148-161	9-11	7	5-6	4	0	2	2
Resultados reales	169	154	10	6	3	2	2	2	2

La previsión que se hace para IU, BNG y CC aporta un cierto valor a la distancia entre el sondeo y los resultados reales, ya que sus valores no coinciden. Sin embargo, si esos tres partidos se hubieran agrupado en una nueva categoría «Otros» (incluyendo también la estimación de 2 que ya existe en esta ca-

peores. También se observa que en algunas elecciones, como las de 1996, unos medios publicaron sondeos muchos más acertados que otros.

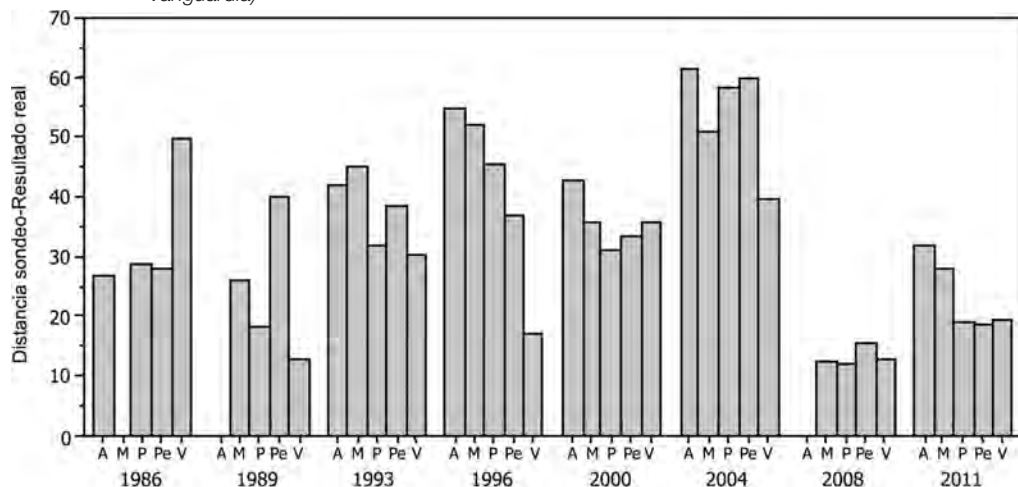
La pregunta que ahora surge es si las distancias observadas son grandes o pequeñas respecto a lo que técnicamente cabría espe-

**TABLA 1.** Distancias entre los sondeos y los resultados finales en las elecciones generales

	1986	1989	1993	1996	2000	2004	2008	2011
ABC	26,8	-	42,1	54,8	42,8	61,6	-	32
<i>El Mundo</i>	-	26	45	52	36	51	12,3	28
<i>El País</i>	28,8	18,2	32	45,7	31	58,2	12	19
<i>El Periódico</i>	28	39,9	38,2	36,8	33,4	59,8	15,5	18,6
<i>La Vanguardia</i>	49,7	13	30,54	17,2	36	39,8	13,1	19,4



**GRÁFICO 1.** Representación gráfica de las distancias entre los sondeos y resultados en las elecciones generales de los años que se indican (A: ABC, M: El Mundo, P: El País, Pe: El Periódico, V: La Vanguardia)



rar. Para poder responderla hemos simulado lo que denominamos «sondeos perfectos».

## UNA DISTANCIA DE REFERENCIA: EL SONDEO PERFECTO

Entendemos por sondeo perfecto aquel que se ha realizado con un tamaño de muestra adecuado y en el que todos los encuestados dicen la verdad: si votarán o si se abstendrán y, en el caso de que voten, a quién lo harán. El mayor tamaño de muestra que aparece en las fichas técnicas de los sondeos publicados es de 15.600 individuos en el que realizó IPSOS-Eco-Consulting para ABC en las elecciones del año 2000. Entendiendo por adecuado un tamaño de muestra grande pero realista se ha tomado una muestra de 20.000 individuos para los sondeos perfectos. Para las elecciones del año 2000 y las siguientes, a cada circunscripción se le ha asignado una proporción de la muestra total igual a la utilizada por el CIS en su encuesta preelectoral<sup>4</sup>. Para las elecciones anteriores al año 2000

no se publican de forma explícita los tamaños de muestra en cada circunscripción y para estas elecciones se han utilizado las mismas proporciones que para el año 2000.

Asignado el tamaño de muestra a cada circunscripción, se ha utilizado la distribución multinomial —con las proporciones realmente obtenidas por cada partido— para simular los resultados que se tendrían en una muestra de ese tamaño en la que todos los encuestados dijieran siempre la verdad. La tabla 2 presenta los resultados reales (proporción de votos y número de escaños) de la provincia de Barcelona en las elecciones de 2011 y la estimación obtenida en una ejecución del «sondeo perfecto». Los votos en la muestra son los que se obtienen por simulación, los votos totales son la extrapolación al total de votos válidos de las proporciones que cada partido ha obtenido en la muestra, y el número de escaños se deduce aplicando la fórmula D'Hondt al total de votos obtenidos.

Sumando los escaños obtenidos en cada circunscripción se tiene una estimación de los resultados generales proporcionada por el son-

<sup>4</sup> La ficha técnica de la encuesta publica el tamaño de muestra diseñada total y el tamaño para cada circunscripción, y de ahí se han deducido las proporciones. La ficha técnica del año 2000 está en: <http://www.cis.es/cis/opencms/-Archivos/Marginales/>

2380\_2399/2382/Ft2382.pdf (consultado el 15 de diciembre de 2011).



**TABLA 2.** *Estimación de los resultados en las elecciones generales de 2011 en la circunscripción de Barcelona mediante un «sondeo perfecto»*

		Resultados reales		Estimación por simulación		
		% Votos	Escaños	Votos en la muestra <sup>(1)</sup>	Votos totales <sup>(2)</sup>	Escaños <sup>(3)</sup>
Abstención		29,35	-	286	-	-
Porcentaje sobre el total de votos válidos	PSC-PSOE	27,78	10	175	498.318	9
	CiU	27,15	9	179	509.708	10
	PP	20,95	7	146	415.739	8
	ICV-EUiA	9,07	3	52	148.072	2
	ERC-RI.cat	6,48	2	46	130.986	2
	PxC	2,02	0	9	25.628	0
	Eb	1,53	0	9	25.628	0
Otros/blanco		5,02	-	29	82.578	-

(1) El tamaño de muestra en la provincia de Barcelona debe ser de 931 individuos, de acuerdo con la proporción aplicada por el CIS en el estudio preelectoral de las elecciones de 2011 (0,04656).

(2) Proporción obtenida en la muestra, extrapolada al total de votos válidos.

(3) Aplicación de la fórmula D'Hondt a los votos totales.

de perfecto. La distancia correspondiente a esa estimación representa un error inevitable, que no se puede eliminar aunque se realice todo el proceso de muestreo de forma perfecta y sin ninguna de las dificultades inherentes a los sondeos electorales comentadas al principio.

Por otra parte, como en los resultados obtenidos por simulación influye el azar, si se repite el proceso no hay que esperar que se obtengan siempre los mismos valores. Para cuantificar la distancia que corresponde al sondeo perfecto se ha repetido su ejecución 10.000 veces y se ha medido la distancia para cada uno de los 10.000 resultados obtenidos. El promedio de esos valores es el que se ha tomado como distancia entre el sondeo perfecto y el resultado final<sup>5</sup>.

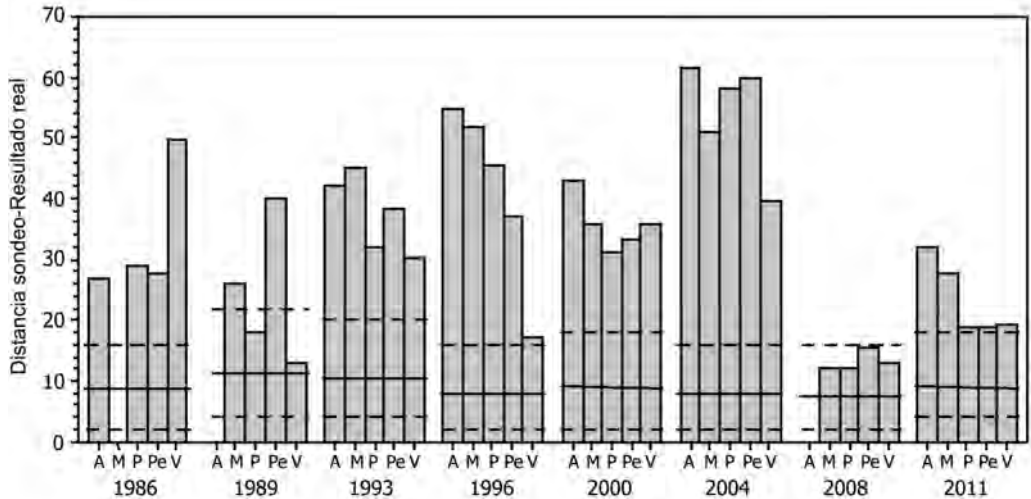
<sup>5</sup> La generación de resultados para los sondeos perfectos y el cálculo de las distancias se ha realizado mediante programas escritos con el paquete de software estadístico R.

En el gráfico 2 se han vuelto a representar las distancias de los sondeos publicados, pero añadiendo una línea horizontal gruesa en el valor correspondiente a la distancia media que presenta el sondeo perfecto para esas elecciones. También se han determinado los percentiles de 97,5 y de 2,5%, que son los valores entre los que se encuentra la distancia del sondeo perfecto en el 95% de los casos. Esos límites están marcados con líneas a trazos.

Puede observarse que en las elecciones de 2008 todos los sondeos publicados presentan una distancia al resultado real que está dentro del intervalo proporcionado por el sondeo perfecto en el 95% de los casos. En otras elecciones alguno de los sondeos está dentro de los límites (1989) o muy cerca de ellos (1996 y 2011), mientras que en el resto (1986, 1993, 2000 y 2004) todos tienen distancias mayores.

El gráfico 3 muestra la distribución de las distancias obtenidas en las 10.000 simula-

**GRÁFICO 2.** Reproducción del gráfico 1 añadiendo en cada elección una línea gruesa horizontal que representa la distancia media obtenida con un sondeo perfecto y las líneas a trazos que marcan la zona en la que aparece esa distancia en el 95% de los casos

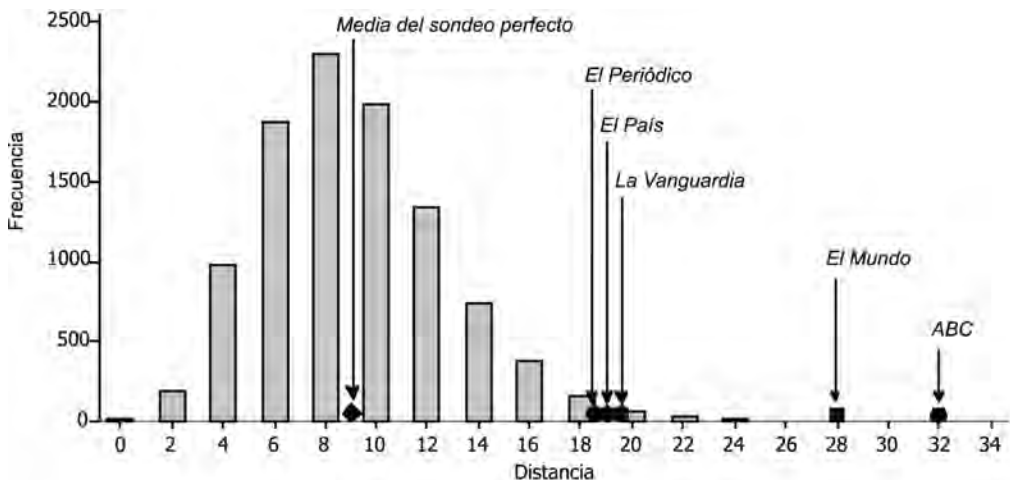


ciones del sondeo perfecto para las elecciones de 2011. En ese mismo gráfico se han situado las distancias correspondientes a los sondeos analizados.

La distribución de las distancias obtenidas a partir de las simulaciones del sondeo

perfecto presenta una forma similar en todas las elecciones, siendo siempre ligeramente asimétricas con cola hacia la derecha. Esta asimetría es más acentuada cuando el valor medio de la distancia es menor, ya que en ese caso tiene más influencia la frontera que

**GRÁFICO 3.** Distribución de las distancias obtenidas en el sondeo perfecto para las elecciones de 2011 y valores correspondientes a los sondeos analizados



significa una distancia igual a cero. En el Anexo 2 se presenta la distribución de las distancias obtenidas para el sondeo perfecto en las ocho elecciones consideradas.

### DIRECCIÓN DEL ERROR EN LA PREDICCIÓN Y TENDENCIA POLÍTICA DEL DIARIO QUE LA PUBLICA

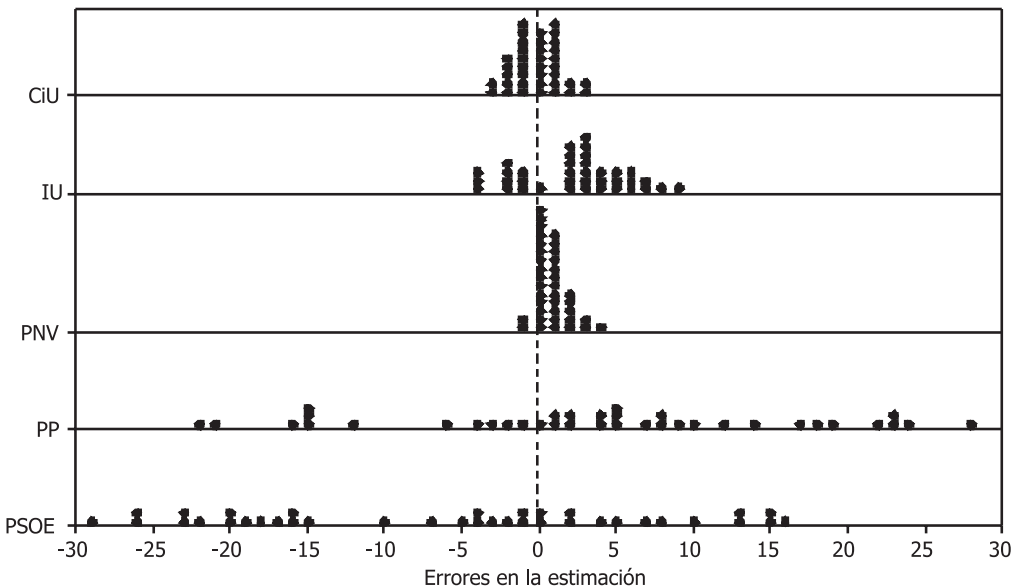
A veces se comenta que los diarios tienden a ser optimistas en sus predicciones sobre los partidos con los que tienen afinidad ideológica y pesimistas sobre los contrarios. Para analizar si los datos avalan esta sospecha se ha definido el error que se comete en la estimación del número de escaños como la diferencia entre el valor medio del intervalo de estimación y el valor real. Por ejemplo, la previsión que publicó el diario *ABC* para el PSOE en las elecciones de 1986 fue un intervalo de 167-180 diputados y el valor real fue de 184, por tanto, el error cometido fue  $(167+180)/2 - 184 = -10,5$ . Nótese que esta

forma de definir el error es equivalente al promedio de las diferencias entre cada valor de los incluidos en el intervalo y el valor real.

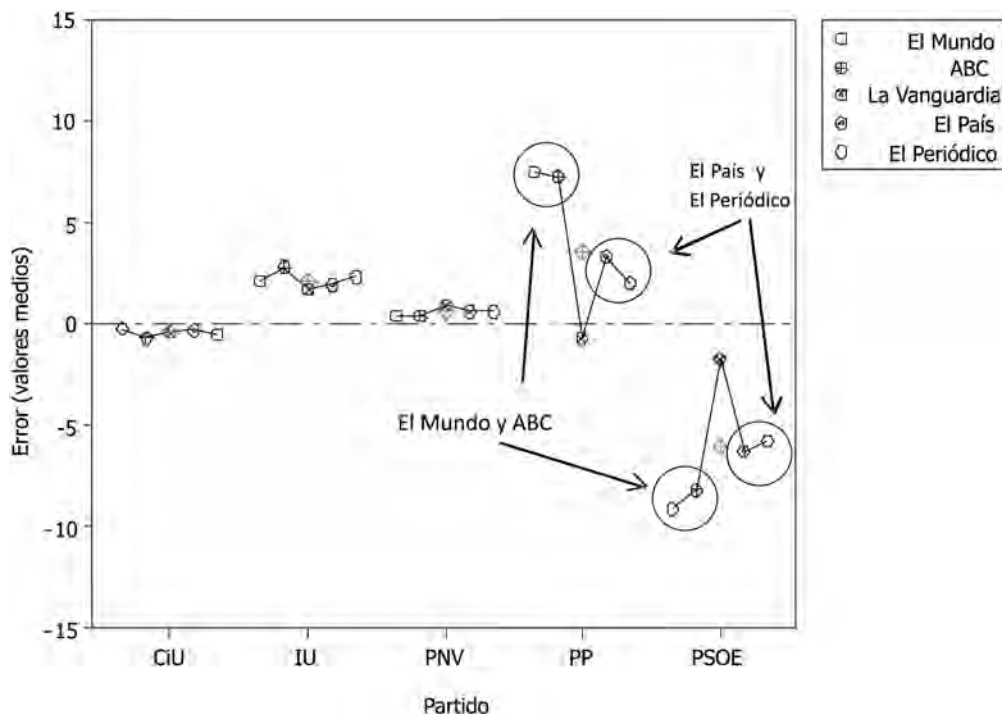
El gráfico 4 muestra los errores cometidos para cada partido agrupando elecciones y diarios (en el Anexo 3 figura la tabla detallada con todos los datos). Salta a la vista que los errores son mayores en los partidos que obtienen un mayor número de escaños (PSOE y PP), tal como prevé la teoría estadística. Para CiU las estimaciones han sido siempre muy buenas, para IU, PNV y PP se ha errado más por exceso que por defecto mientras que para el PSOE se observa una mayor frecuencia de errores negativos.

Los errores medios, separando las estimaciones por diarios pero manteniendo agrupadas todas las elecciones (las medias son de las 8 elecciones consideradas) aparecen en el gráfico 5. Se observa que tanto el acierto para CiU y PNV como la sobreesimación para IU se da de forma muy similar en los cinco diarios. Asimismo se pone de

**GRÁFICO 4.** Errores en la estimación de los escaños de cada partido agrupando diarios y elecciones



**GRÁFICO 5.** Errores medios en la estimación del número de escaños obtenido por cada partido separando los cinco diarios considerados



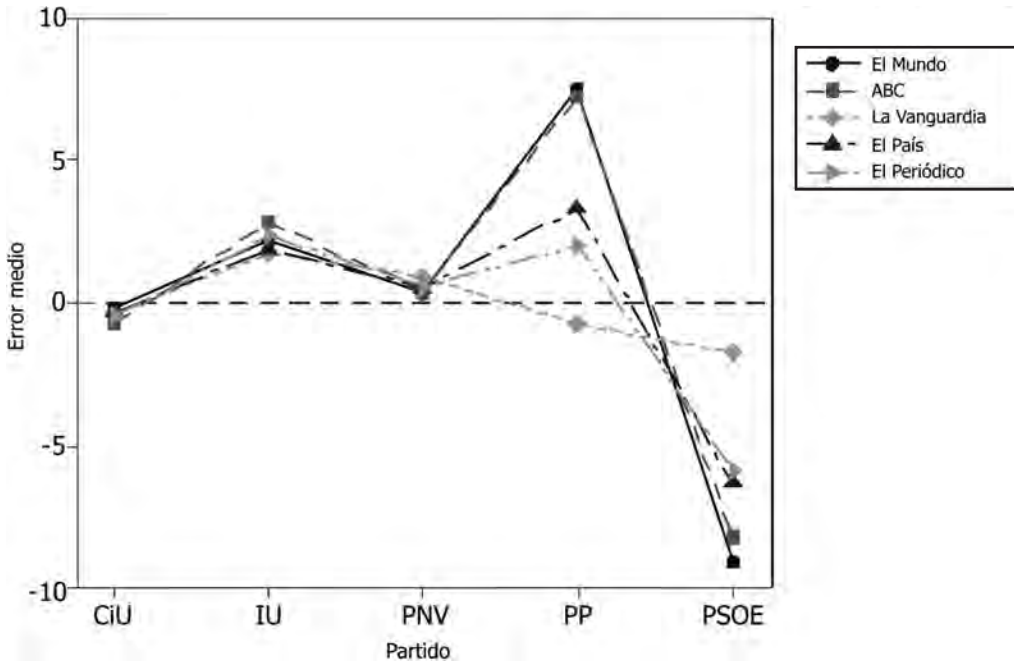
manifiesto la mayor variabilidad entre diarios que presentan las estimaciones para el PP y PSOE. También se observa que *ABC* y *El Mundo* sobreestiman más al PP e infraestiman más al PSOE, mientras que *El Periódico* y *El País* hacen justo lo contrario.

Los gráficos 5 y 6 son similares, pero en este último los puntos correspondientes a cada partido están sobre la misma vertical y los de cada diario están unidos por una línea. Se observa que para CiU, IU y PNV los cinco diarios hacen un recorrido similar, mientras que para PP y PSOE las líneas difieren claramente: por un lado van (casi superpuestas) las que corresponden a *El Mundo* y *ABC* (los diarios que más sobreestiman los resultados del PP y los que más infraestiman los del PSOE); por otro lado tenemos —también con líneas muy próximas— a *El País* y *El Periódico* (los diarios que sobreestiman más al PSOE y que más infraestiman al PP). *La Vanguardia* presenta una trayectoria en solitario.

Aunque la cercanía diario-partido puede ser discutible, en general hay consenso en considerar que *ABC* y *El Mundo* son más afines al PP<sup>6</sup> mientras que *El País* y *El Periódico* lo son al PSOE. *La Vanguardia* no es clasificable en este terreno. Así pues, el análisis gráfico está en línea con la idea de que los diarios tienden a ser más optimistas con los partidos a los que son más cercanos. Sin embargo, el análisis de la varianza de los errores no muestra que la interacción diario-partido sea estadísticamente significativa, es decir, que el error cometido en la estimación del número de escaños que ob-

<sup>6</sup> Castromil (2012) realiza un detallado estudio sobre la publicación en *ABC*, *El Mundo* y *El País* de noticias que pueden perjudicar a un determinado partido durante la campaña electoral y llega a la conclusión de que los diarios que publican más noticias negativas para el PSOE son *ABC* y *El Mundo* (no existe fiabilidad estadística para otros partidos).

**GRÁFICO 6.** Errores medios (en las 8 elecciones consideradas) en la estimación del número de escaños según el medio que los publica



tendrá un partido no depende del diario que lo publica de forma estadísticamente significativa. Pero también es verdad que si se realiza el análisis sin contar los resultados de las últimas elecciones en las que *ABC* sobreestimó mucho al PSOE y *El País* al PP (al contrario que en la mayoría de las otras elecciones), la diferencia sí que aparece claramente significativa. Estamos, por tanto, ante una situación sobre la que no pueden realizarse afirmaciones contundentes. El Anexo 3 contiene los análisis estadísticos detallados.

### ALGUNAS CONSIDERACIONES SOBRE LA FICHA TÉCNICA QUE ACOMPAÑA A LA PUBLICACIÓN DE LOS SONDEOS

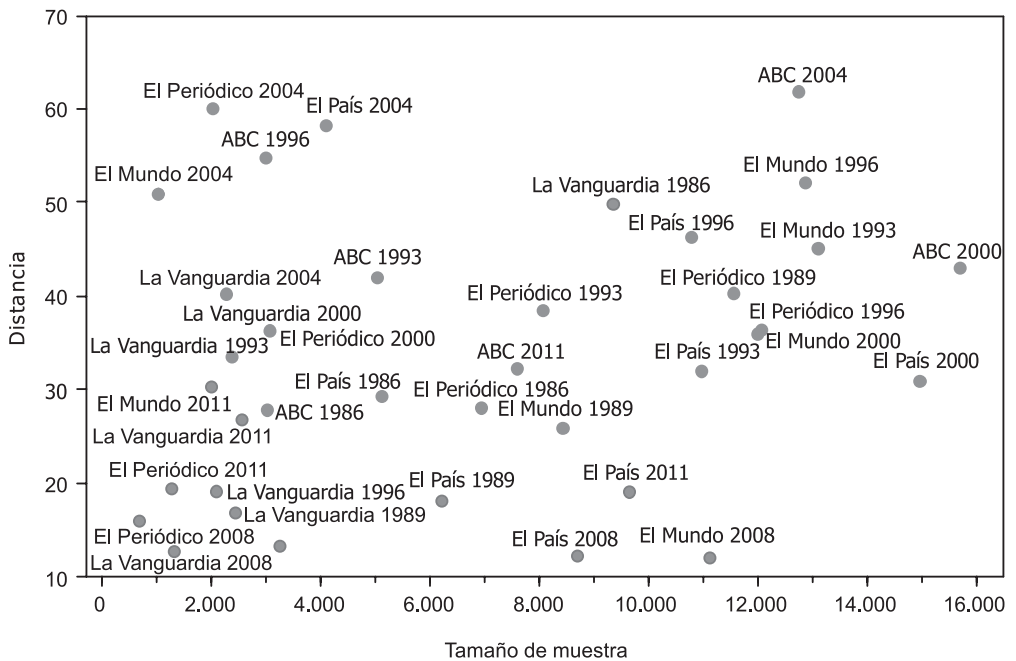
Desde la publicación de la Ley 14/1980, de 18 de abril, sobre régimen de encuestas

electorales<sup>7</sup>, todos los sondeos que se publican incluyen una «ficha técnica» donde se informa, con más o menos detalle, sobre el procedimiento seguido para obtener la muestra, su tamaño y el margen de error para un determinado nivel de confianza (bajo ciertos supuestos, como es habitual).

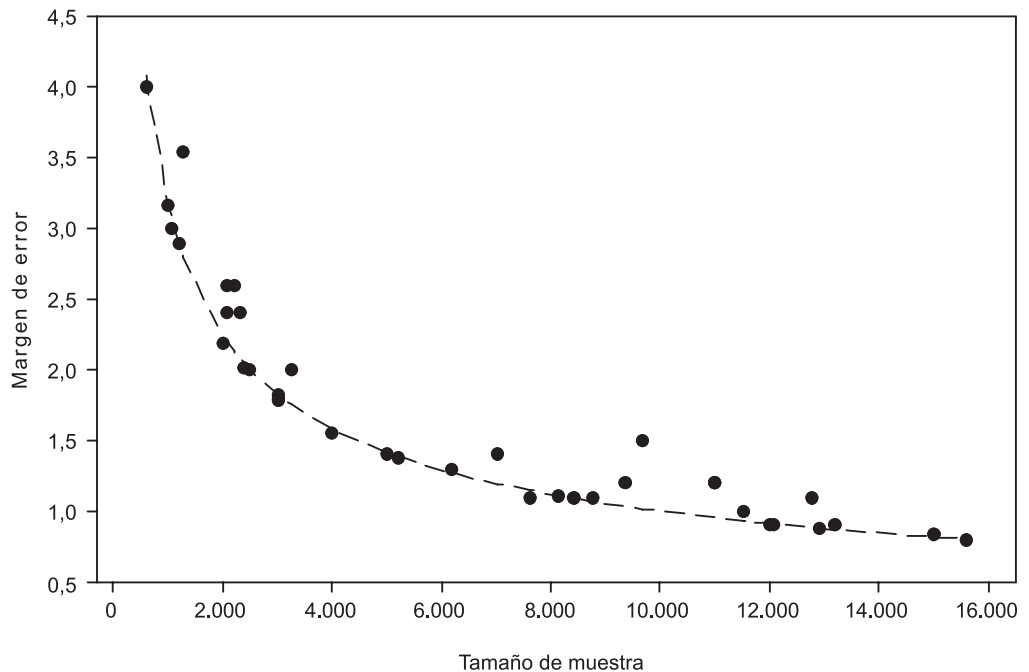
Pero el interés de la información que aportan es discutible. Por ejemplo, cuando se informa de un mayor tamaño de muestra (mayor esfuerzo, más presupuesto) cabe esperar una mayor calidad en la estimación, pero el análisis de los datos no avala este hecho. El gráfico 7 representa la distancia del sondeo al resultado real en función del tamaño de la muestra utilizada.

<sup>7</sup> Un detallado análisis de los antecedentes y del contenido de esta ley puede verse en Gálvez (2000).

**GRÁFICO 7.** *Relación entre distancia al resultado real y tamaño de muestra en los sondeos considerados*



**GRÁFICO 8.** *Relación entre el margen de error y el tamaño de muestra que consta en la ficha técnica de cada sondeo. La línea a trazos corresponde a la relación para un muestreo aleatorio simple, con  $p = q = 0,5$*



Lo razonable sería observar una correlación negativa, es decir, a mayor tamaño de muestra menor distancia al resultado real, pero dicha correlación no existe: el coeficiente de correlación tiene un valor de 0,2 (positivo) con un p-valor de 0,272.

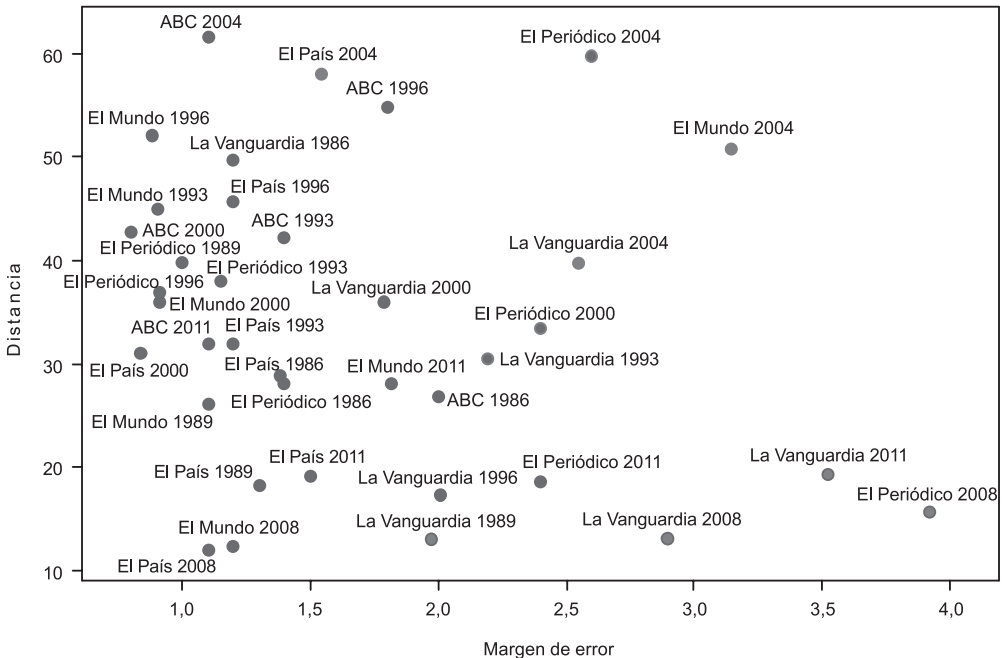
Sí existe relación entre el tamaño de muestra y margen de error que se indica en la ficha técnica porque uno se calcula a partir del otro. El gráfico 8 muestra el margen de error en función del tamaño de muestra para cada uno de los sondeos considerados. También se ha representado la relación matemática que existe entre ambas variables en el supuesto de muestreo aleatorio simple, población infinita y  $p = q = 0,5$ . La similitud entre la tendencia que marcan los puntos que representan los sondeos y la línea que corresponde al muestreo aleatorio simple es muy clara, lo que conduce a pensar que en muchos casos se ha utilizado esa fórmula (solo válida para el caso del

muestreo aleatorio simple) con independencia de la estrategia de muestreo que se haya seguido.

Tampoco existe relación entre la calidad de la estimación y el margen de error que se indica en la ficha técnica (gráfico 9). El coeficiente de correlación entre ambas variables es -0,236 (aquí es negativo cuando cabría esperar que fuera positivo) con un p-valor de 0,160.

Por tanto, los datos demuestran que la información técnica (tamaño de muestra, margen de error, nivel de confianza...) que acompaña a los sondeos no es útil para valorar la calidad de sus previsiones, tal como pretende la ley que obliga a su publicación. Seguramente lo más relevante es el nombre de la empresa que lo ha realizado y su nivel de acierto en sondeos anteriores. Este nivel de acierto puede ser medido a través de un solo número, fácil de interpretar y sencillo de calcular, como la distancia sondeo-realidad usada en este trabajo.

**GRÁFICO 9.** Relación entre distancia al resultado real y margen de error que figura en la ficha técnica de los sondeos considerados





## CONCLUSIONES

A pesar de las dificultades que entraña la realización de un sondeo electoral —dificultades que van más allá de lo que es capaz de resolver la teoría estadística—, los sondeos publicados en España una semana antes de las elecciones generales presentan, salvo en algunos casos, un nivel de acierto que puede considerarse razonablemente bueno.

A esta conclusión se ha llegado después de definir una distancia entre las estimaciones realizadas y el resultado real de las elecciones, medir esa distancia para los sondeos publicados y compararla con la que se obtiene en un «sondeo perfecto», que es un sondeo realizado por simulación con un tamaño de muestra grande (mayor que cualquiera de los sondeos publicados) y utilizando los resultados reales para generar la opción elegida por cada uno de los individuos de la muestra. Existe una clara diferencia en las distancias cuando se comparan las distintas elecciones. Los sondeos que presentan peores resultados son los de las elecciones de 2004 (posteriores al atentado del 11-M) y los mejores son los de las de 2008, en las que todos los sondeos publicados presentan una distancia que entra dentro del intervalo que cabe esperar en un sondeo perfecto. También las diferencias son claras entre partidos: el que mejor se ha estimado es CiU y los más difíciles son los que obtienen mayor número de diputados (PSOE y PP), como era de esperar.

El análisis gráfico de los datos confirma la idea de que los diarios son más optimistas en sus previsiones sobre los resultados de los partidos a los que son afines, pero la interacción diario-partido no aparece como significativa realizando un estudio más detallado a través del análisis de la varianza.

Finalmente, se ha constatado que el contenido de la ficha técnica parece puramente formal, ya que no aporta ninguna información

que sirva para valorar el nivel de calidad, o de fiabilidad, de los sondeos que se presentan. Seguramente la forma de elegir la muestra y de «cocinar» los datos (dicho sea en el mejor sentido de la palabra: asignar los votos a los indecisos, etc.) forma parte del *know how* de la empresa que realiza el sondeo. Ahí es donde está el secreto de la buena estimación, y eso no consta en la ficha técnica.

## BIBLIOGRAFÍA

- Bouza, Fermín (1998). «Comunicación política: encuestas, agendas y procesos cognitivos electorales». *Praxis Sociológica*, 3: 49-58.
- Castromil, Antón R. (2012). «Negativismo mediático y campaña electoral en las Elecciones Generales de 2008». *Revista Española de Investigaciones Sociológicas*, 139: 163-174.
- Delicado, Pedro y Frederic Udina (2001). «¿Cómo y cuánto fallan los sondeos electorales?». *Revista Española de Investigaciones Sociológicas*, 96: 123-150.
- Díaz de Rada, Vidal (2007). «El trabajo de campo con encuesta presencial. Algunas reflexiones derivadas de la práctica profesional». *Investigación y Marketing*, 96: 11-26.
- Gálvez, Luis (2000). «Organismos de sondeos, encuestas electorales y derecho». *Revista de Estudios Políticos*, 110: 97-121.
- Langsrud, Øyvind (2003). «ANOVA for unbalanced data: Use Type II instead of Type III sums of squares». *Statistics and Computing*, 13: 163-167.
- Manly, Bryan F. J. (1986). *Multivariate Statistical Methods. A primer*. Boca Raton: Chapman & Hall/CRC.
- Pavía, José M. et al. (2011). «Encuestas electorales online: nuevos retos, viejos problemas». *Revista Española de Investigaciones Sociológicas*, 135: 107-122.
- y Beatriz Larráz (2012). «Sesgo de no-respuesta y modelos de superpoblación en encuestas electorales». *Revista Española de Investigaciones Sociológicas*, 137: 121-150.
- Udina, Frederic y Pedro Delicado (2005). «Estimating Parliamentary composition through electoral polls». *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168: 387-399.

**RECEPCIÓN:** 07/02/2012

**REVISIÓN:** 16/11/2012

**APROBACIÓN:** 08/02/2013



**ANEXO 1.** *Sondeos publicados en los medios considerados y resultados reales en las elecciones al Congreso de los Diputados de 1986 a 2011*  
*La última columna indica la fecha y la página del diario donde se publicó el sondeo.*

Elecciones del 22 de junio de 1986

Sondeos / Resultados	PSOE	AP-PDP-PL	CDS	CiU	IU	EAJ-PNV	HB	EE	Otros	Fecha (Pág.)
ABC	167-180	105-121	13-19	17-20	6-12	8-10	3	1	3	16/06 (20)
<i>El Mundo</i> (1)	-	-	-	-	-	-	-	-	-	-
<i>El País</i>	167-194	86-105	15-29	17-20	10-16	8-9	3-4	1-2	4-7	15/06 (11)
<i>El Periódico</i>	180	98	30	17	5	7	4	1	8	15/06 (3)
<i>La Vanguardia</i>	178-199	71-90	17-28	20-21	7-12	9-11	1-2	1-2	11-17	16/06 (5)
<b>Result. reales</b>	<b>184</b>	<b>105</b>	<b>19</b>	<b>18</b>	<b>7</b>	<b>6</b>	<b>5</b>	<b>2</b>	<b>4</b>	

Elecciones del 29 de octubre de 1989

Sondeos / Resultados	PSOE	PP	CiU	IU	CDS	EAJ-PNV	HB	PA	Otros	Fecha (Pág.)
ABC (2)	-	-	-	-	-	-	-	-	-	-
<i>El Mundo</i>	168	103	17	16	23	5	5	2	11	23/10 (1)
<i>El País</i>	176-182	95-104	18	15-16	16-19	5-7	4-5	2-3	6-8	22/10 (13)
<i>El Periódico</i>	172-178	87-91	17-19	22-26	18-21	7	7	4	7-10	22/10 (3)
<i>La Vanguardia</i>	172-178	104-107	17-18	13	14-15	7-8	3	3-4	5-12	23/10 (1)
<b>Result. reales</b>	<b>175</b>	<b>107</b>	<b>18</b>	<b>17</b>	<b>14</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>8</b>	

Elecciones del 6 de junio de 1993

Sondeos / Resultados	PSOE	PP	IU	CiU	EAJ-PNV	CC	HB	ERC	Otros	Fecha (Pág.)
ABC	134-144	144-153	26-28	16-17	6-7	3-4	3	1	5-7	29/05 (23)
<i>El Mundo</i>	132-141	147-155	25-27	17-19	5-6	3	3-4	1	5-6	30/05 (7)
<i>El País</i>	135-151	141-155	20-22	18-20	6	3	2-3	1	5-8	30/05 (13)
<i>El Periódico</i>	133-152	134-151	19-24	19-21	6-7	4-6	3-4	1	6-13	30/05 (4)
<i>La Vanguardia</i>	142-146	144-148	24-25	18-19	6-7	*	*	0-1	9-11	30/05 (18)
<b>Result. reales</b>	<b>159</b>	<b>141</b>	<b>18</b>	<b>17</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>1</b>	<b>3</b>	

Elecciones del 3 de marzo de 1996

Sondeos / Resultados	PP	PSOE	IU	CiU	EAJ-PNV	CC	BNG	HB	Otros	Fecha (Pág.)
ABC	176-184	117-125	22-28	13-14	4-5	*	*	*	6-8	25/02 (5)
<i>El Mundo</i>	170-179	113-123	25-29	14-15	6-7	2-3	2	2	3	25/02 (9)
<i>El País</i>	170-178	118-128	24-27	13-15	5-6	3	1	2	2-3	25/02 (16)
<i>El Periódico</i>	165-175	119-129	18-27	14-16	5	2-3	1	2	2-5	20/02 (5)
<i>La Vanguardia</i>	160-170	135-145	19-21	14-15	5-6	*	*	*	7-8	25/02 (13)
<b>Result. reales</b>	<b>156</b>	<b>141</b>	<b>21</b>	<b>16</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>2</b>	<b>3</b>	

(1) En esta fecha el diario *El Mundo* todavía no había salido.

(2) Para las elecciones de 1989 el diario ABC no publicó ningún sondeo.

\* Incluido en Otros.

**ANEXO 1.** Sondeos publicados en los medios considerados y resultados reales en las elecciones al Congreso de los Diputados de 1986 a 2011 (*continuación*)

Elecciones del 12 de marzo de 2000

Sondeos / Resultados	PP	PSOE	CiU	IU	EAJ-PNV	CC	BNG	PA	Otros	Fecha (Pág.)
<i>ABC</i>	158-164	138-144	15-16	9-12	6-7	4	3-5	*	3-4	05/03 (1)
<i>El Mundo</i>	164-170	137-143	16	9-11	6-7	4	3-5	0-1	2	05/03 (9)
<i>El País</i>	165-171	131-139	14-15	9-13	6-7	4	4-5	0	3-5	05/03 (17)
<i>El Periódico</i>	165-170	135-140	13-14	9-12	7-8	4-5	4-5	2	1	05/03 (4)
<i>La Vanguardia</i>	166-170	136-140	14-15	12-14	6	4	3	*	2-5	05/03 (1)
<b>Result. reales</b>	<b>183</b>	<b>125</b>	<b>15</b>	<b>8</b>	<b>7</b>	<b>4</b>	<b>3</b>	<b>1</b>	<b>4</b>	

Elecciones del 14 de marzo de 2004

Sondeos / Resultados	PSOE	PP	CiU	ERC	EAJ-PNV	IU	CC	BNG	Otros	Fecha (Pág.)
<i>ABC</i>	133-137	174-177	10-11	6-7	7	8-10	3	2	2	29/02 (1)
<i>El Mundo</i>	138-144	168-173	11	5-6	7	6-8	3-4	3	1-2	08/03 (8)
<i>El País</i>	134-141	168-172	10-11	7	7-8	11	3-4	3	2	07/03 (14)
<i>El Periódico</i>	135-140	169-173	10-11	6-7	7	10-12	*	*	10-14	07/03 (3)
<i>La Vanguardia</i>	143-147	162-167	10	5-6	7	9-10	*	*	4-5	07/03 (1)
<b>Result. reales</b>	<b>164</b>	<b>148</b>	<b>10</b>	<b>8</b>	<b>7</b>	<b>5</b>	<b>3</b>	<b>2</b>	<b>3</b>	

Elecciones del 9 de marzo de 2008

Sondeos / Resultados	PSOE	PP	CiU	EAJ-PNV	ERC	IU	BNG	CC	Otros	Fecha (Pág.)
<i>ABC (3)</i>	-	-	-	-	-	-	-	-	-	02/03 (14)
<i>El Mundo</i>	157-171	148-161	9-11	7	5-6	4	0	2	2	03/03 (10)
<i>El País</i>	165-169	148-154	9	7	5-6	4	2	2	2-3	02/03 (14)
<i>El Periódico</i>	165-170	149-154	8-9	6	5-6	4-5	*	*	8-11	28/02 (18)
<i>La Vanguardia</i>	162-167	152-156	8-9	6-7	6	4-5	*	*	5-7	03/03 (14)
<b>Result. reales</b>	<b>169</b>	<b>154</b>	<b>10</b>	<b>6</b>	<b>3</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	

Elecciones del 20 de noviembre de 2011

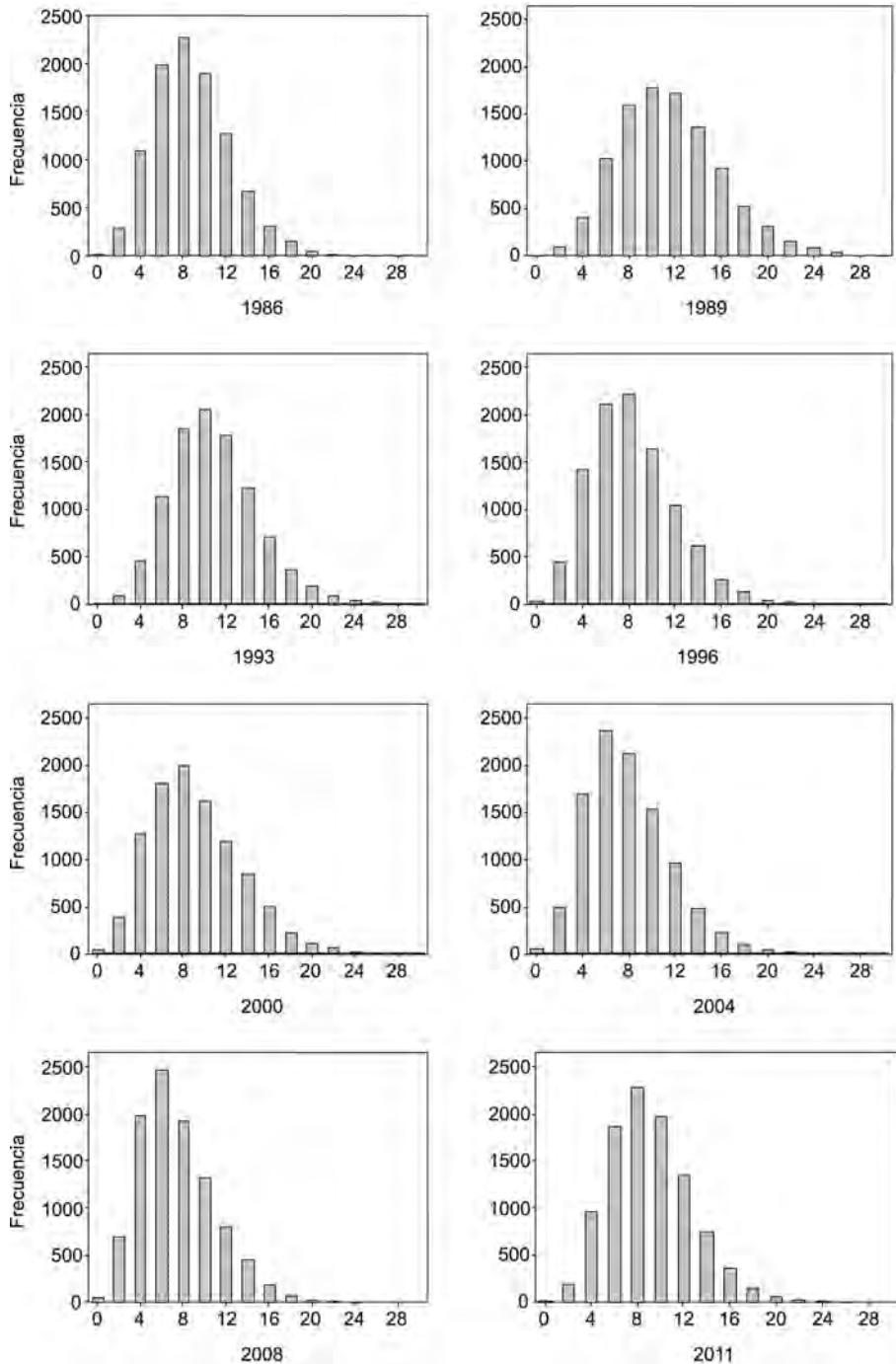
Sondeos/ Resultados	PP	PSOE	CiU	IU	AMAIUR	UPyD	EAJ-PNV	ERC	Otros	Fecha (Pág.)
<i>ABC</i>	187-188	123-126	12-14	6-7	5-6	2-3	4	2	5-6	13/11 (19)
<i>El Mundo</i>	198	112	14	7	3	3	5	2	6	13/11 (4)
<i>El País</i>	192-196	110-113	14	11	4-5	2	5	3	5	13/11 (14)
<i>El Periódico</i>	188-140	115-118	13-14	8-10	4-5	4	4-5	2-3	7-8	13/11 (31)
<i>La Vanguardia</i>	184-147	116-120	12-14	8-10	3-4	3-4	4-5	3	8-9	13/11 (17)
<b>Result. reales</b>	<b>186</b>	<b>110</b>	<b>16</b>	<b>11</b>	<b>7</b>	<b>5</b>	<b>5</b>	<b>3</b>	<b>7</b>	

(3) Da las previsiones en porcentajes, pero no en número de escaños.

\* Incluido en *Otros*.

**ANEXO 2.** Distribución de los valores de la distancia en las simulaciones del sondeo perfecto

**GRÁFICO A2.1.** Distribución de las distancias obtenidas repitiendo 10.000 veces el sondeo perfecto en cada una de las elecciones consideradas



**ANEXO 3.** Tabla de errores en cada partido (diferencia entre el valor medio del intervalo de predicción y el valor real) en los sondeos considerados

		PSOE	PP*	CiU	IU	PNV
1986	ABC	-10,5	5	0,5	2	3
	<i>El Mundo</i>	-	-	-	-	-
	<i>El País</i>	-3,5	-6,5	0,5	-2	1
	<i>El Periódico</i>	-4	4	-1	-2	1
	<i>La Vanguardia</i>	4,5	-21	2,5	2,5	4
1989	ABC	-	-	-	-	-
	<i>El Mundo</i>	-7	5	-1	-1	0
	<i>El País</i>	4	-4	0	-1,5	1
	<i>El Periódico</i>	0	-12,5	0	7	2
	<i>La Vanguardia</i>	0	-1	-0,5	-4	2,5
1993	ABC	-20	7,5	-0,5	9	1,5
	<i>El Mundo</i>	-22,5	10	1	8	0,5
	<i>El País</i>	-16	7	2	3	1
	<i>El Periódico</i>	-16,5	1,5	3	3,5	1,5
	<i>La Vanguardia</i>	-15	5	1,5	6,5	1,5
1996	ABC	-20	24	-2,5	4	-0,5
	<i>El Mundo</i>	-23	18,5	-1,5	6	1,5
	<i>El País</i>	-18	18	-2	4,5	0,5
	<i>El Periódico</i>	-17	14	-1	1,5	0
	<i>La Vanguardia</i>	-1	9	-1,5	-1	0,5
2000	ABC	16	-22	0,5	2,5	-0,5
	<i>El Mundo</i>	15	-16	1	2	-0,5
	<i>El País</i>	10	-15	-0,5	3	-0,5
	<i>El Periódico</i>	12,5	-15,5	-1,5	2,5	0,5
	<i>La Vanguardia</i>	13	-15	-0,5	5	-1
2004	ABC	-29	27,5	0,5	4	0
	<i>El Mundo</i>	-23	22,5	1	2	0
	<i>El País</i>	-26,5	22	0,5	6	0,5
	<i>El Periódico</i>	-26,5	23	0,5	6	0
	<i>La Vanguardia</i>	-19	16,5	0	4,5	0
2008	ABC	-	-	-	-	-
	<i>El Mundo</i>	-5	0,5	0	2	1
	<i>El País</i>	-2	-3	-1	2	1
	<i>El Periódico</i>	-1,5	-2,5	-1,5	2,5	0
	<i>La Vanguardia</i>	-4,5	0	-1,5	2,5	0,5
2011	ABC	14,5	1,5	-3	-4,5	-1
	<i>El Mundo</i>	2	12	-2	-4	0
	<i>El País</i>	1,5	8	-2	0	0
	<i>El Periódico</i>	6,5	4	-2,5	-2	-0,5
	<i>La Vanguardia</i>	8	0,5	-3	-2	-0,5

\* En las elecciones de 1986 los escaños asignados al PP son los de la coalición AP-PDP-PL junto con los que obtuvo el CDS. En las elecciones de 1989 son los del PP más los del CDS.

La existencia de valores faltantes (*El Mundo* en 1986 y *ABC* en 1989 y 2008) dificulta el análisis de los datos para determinar la significación estadística de la interacción diario-partido a través del análisis de la varianza. Una forma razonable de resolver esta dificultad es asignar a cada valor faltante el promedio de los valores existentes para ese mismo diario y partido. Así, por ejemplo, el valor faltante que corresponde al error cometido por *El Mundo* en la estimación del número de escaños obtenidos por el PSOE en las elecciones de 1986 ha sido sustituido por el promedio de los otros errores cometidos también por *El Mundo* con el PSOE, es decir: -7 (en 1989), -22,5 (1993), -23 (1996), 15 (2000), -23 (2004), -5 (2008) y 2 (2011), lo cual da un valor de -9,07. De igual forma se ha procedido con el resto de valores faltantes<sup>1</sup>.

Resuelto el problema de los valores faltantes se ha centrado el estudio en los errores cometidos con PSOE y PP por los diarios considerados afines a uno u otro partido. En el análisis de la varianza se ha utilizado un modelo de efectos fijos con los tres factores: Año, Diario y Partido y las tres interacciones: Año-Diario, Año-Partido y Diario-Partido, siendo esta última la que realmente interesa. El resultado obtenido con el paquete de software estadístico MINITAB es el siguiente:

General Linear Model: Error versus Año; Diario; Partido						
Factor	Type	Levels	Values			
Año	fixed	8	1986; 1989; 1993; 1996; 2000; 2004; 2008; 2011			
Diario	fixed	4	ABC; El Mundo; El País; El Periódico			
Partido	fixed	2	PP; PSOE			

Analysis of Variance for Error, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Año	7	660,41	660,41	94,34	3,13	0,020
Diario	3	20,87	20,87	6,96	0,23	0,874
Partido	1	2442,95	2442,95	2442,95	81,14	0,000
Año*Diario	21	121,29	121,29	5,78	0,19	1,000
Año*Partido	7	8822,33	8822,33	1260,33	41,86	0,000
Diario*Partido	3	220,97	220,97	73,66	2,45	0,092 ←
Error	21	632,30	632,30	30,11		
Total	63	12921,12				

S = 5,48722R-Sq = 95,11%R-Sq(adj) = 85,32%

<sup>1</sup> También puede resolverse calculando la tabla de análisis de la varianza usando las sumas de cuadrados de Tipo II, tal y como se recomienda en Langsrud (2003). Lo hemos realizado con el software estadístico R usando la función Anova del paquete «car» y los resultados son prácticamente los mismos.

La interacción Diario-Partido aparece con un p-valor de 0,092 y, por tanto, con el criterio habitual de tomar un nivel de significación del 5%, no puede considerarse estadísticamente significativa. Sin embargo, excluyendo del estudio los resultados de las últimas elecciones de noviembre de 2011, sí aparece como significativa:

General Linear Model: Error (sin 2011) versus Año; Diario; Partido

Factor	Type	Levels	Values
Año	fixed	7	1986; 1989; 1993; 1996; 2000; 2004; 2008
Diario	fixed	4	ABC; El Mundo; El País; El Periódico
Partido	fixed	2	PP; PSOE

Analysis of Variance for Error, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Año	6	158,03	158,03	26,34	1,14	0,378
Diario	3	13,18	13,18	4,39	0,19	0,901
Partido	1	2777,84	2777,84	2777,84	120,53	0,000
Año*Diario	18	115,23	115,23	6,40	0,28	0,995
Año*Partido	6	8487,32	8487,32	1414,55	61,38	0,000
Diario*Partido	3	279,82	279,82	93,27	4,05	0,023 ←
Error	18	414,83	414,83	23,05		
Total	55	12246,25				

S = 4,80064    R-Sq = 96,61%    R-Sq(adj) = 89,65%

**ANEXO 4.** Información obtenida de las fichas técnicas de los sondeos analizados

		Empresa que lo realiza	Tamaño de muestra	Margen de error	Fecha y página de publicación
1986	<i>ABC</i>	Aresco, SA	2.505	2,00	16/06 p. 21
	<i>El Mundo</i>	–	–	–	–
	<i>El País</i>	Inst. Demoscopia	5.200	1,38	15/06 p. 16
	<i>El Periódico</i>	IOPE-ETMAR	7.029	1,40	15/06 p. 4
	<i>La Vanguardia</i>	Line Staff	9.371	1,20	16/06 p. 15
1989	<i>ABC</i>	–	–	–	–
	<i>El Mundo</i>	Sigma Dos	8.400	1,10	23/10 p. E12
	<i>El País</i>	Inst. Demoscopia	6.200	1,30	22/10 p. 18
	<i>El Periódico</i>	ICP/Research	11.524	1,00	22/10 p. 4
	<i>La Vanguardia</i>	Inst. Opina	3.262	2,00	23/10 p. 13
1993	<i>ABC</i>	Gruppo, SA	5.000	1,40	29/05 p. 23
	<i>El Mundo</i>	Sigma Dos	13.200	0,90	30/05 p. 13
	<i>El País</i>	Demoscopia SA	11.000	1,20	30/05 p. 14
	<i>El Periódico</i>	Vox Publica / IOPE Etmar	8.148	1,11	30/05 p. 4
	<i>La Vanguardia</i>	Inst. Opina	2.000	2,19	30/05 p. 21
1996	<i>ABC</i>	Tabula V	3.000	1,80	25/02 p. 29
	<i>El Mundo</i>	Sigma Dos & Vox Publica	12.900	0,88	25/02 p. 15
	<i>El País</i>	Demoscopia	11.000	1,20	25/02 p. 17
	<i>El Periódico</i>	Vox Publica	12.069	0,91	20/02 p. 6
	<i>La Vanguardia</i>	Inst. Opina	2.369	2,01	25/02 p. 17
2000	<i>ABC</i>	Ipsos - Ecoconsulting	15.600	0,80	05/03 p. 21
	<i>El Mundo</i>	Sigma Dos	12.000	0,91	05/03 p. 9
	<i>El País</i>	Demoscopia	15.000	0,84	05/03 p. 19
	<i>El Periódico</i>	Vox Publica	2.300	2,40	05/03 p. 6
	<i>La Vanguardia</i>	Inst. Opina	3.000	1,79	05/03 p. 20
2004	<i>ABC</i>	TSN-Demoscopia	12.760	1,10	29/02 p. 15
	<i>El Mundo</i>	Sigma Dos	1.000	3,16	08/03 p. 8
	<i>El País</i>	Inst. Opina	4.000	1,55	07/03 p. 15
	<i>El Periódico</i>	Vox Publica	2.071	2,60	07/03 p. 9
	<i>La Vanguardia</i>	Inst. Noxa	2.200	2,60	07/03 p. 16
2008	<i>ABC</i>	Inst. DYM	1.075	3,00	02/03 p. 14
	<i>El Mundo</i>	Sigma Dos	11.000	1,20	03/03 p. 11
	<i>El País</i>	Metroscopia	8.750	1,10	02/03 p. 14
	<i>El Periódico</i>	GESOP	600	4,00	28/02 p. 18
	<i>La Vanguardia</i>	Inst. Noxa	1.200	2,89	03/03 p. 16
2011	<i>ABC</i>	Inst. DYM	7.612	1,10	02/03 p. 14
	<i>El Mundo</i>	Sigma Dos	3.000	1,82	03/03 p. 11
	<i>El País</i>	Metroscopia	9.675	1,50	02/03 p. 14
	<i>El Periódico</i>	GESOP	2.070	2,40	28/02 p. 18
	<i>La Vanguardia</i>	Inst. Noxa	1.275	3,54	03/03 p. 16