

AT A TIME OF INSURGENT PARTIES, CAN SOCIETIES BELIEVE IN ELECTION POLLS? THE SPANISH EXPERIENCE

JOSÉ I. CASTILLO-MANZANO

LOURDES LÓPEZ-VALPUESTA

RAFAEL POZO-BARAJAS

*Applied Economics & Management Research Group
Universidad de Sevilla*

The main purpose of this paper is to use the Spanish case, through an econometric analysis of 226 electoral polls, to explain why polls are making more mistakes in times of great socioeconomic slumps, political instability and the emergence of new political parties. In this context, it is the very instrument with which society tries to reduce the reigning uncertainty that, paradoxically, can ultimately drive uncertainty up. Our results show that the prediction error for the new emerging parties is significantly higher than for the traditional parties and this error is not sensitive to solutions for increasing the reliability of surveys, such as increasing sample size, transparency constantly conducting periodical surveys, the closeness of the approaching election or the survey mode that is used. It can be observed that pollsters do not want to make predictions that vary greatly from the average of the other polls. Finally, editorial bias appears to play a significant role, especially in the case of traditional parties.

Key words: election polls, insurgent parties, inaccuracy, Spain,

JEL classification: D7.

According to Castillo-Manzano *et al.* (2017), a pattern of electoral volatility and transformation can be observed in the European countries that have suffered most from the social consequences of the economic crisis and required EU intervention, such as Greece, Portugal and Spain [Camacho *et al.* (2018), García Serrano, (2011)]. First, a change of government, with the party that was in opposition at the beginning of the crisis coming out on top. Then, during the second phase, when the economic crisis did not abate, disaffection with politics, or anti-politics [Clarke, (2015)], sprang up, triggering the emergence of new political parties that have variously been defined as European insurgent (ECFR, 2016) or populist [Hameleers *et al.* (2017)] parties.

In line with the above, during the first phase in 2011, in the wake of the strict economic adjustment plan passed by the socialist government (PSOE) in 2010, the

People's Party (PP), which had up until that moment been the leading opposition party, gained an absolute majority in the Congress (lower house) [Chari, (2013)].

However, the turning point in Spanish political life in recent years was marked by the Spanish elections to the European Parliament in 2014. The pre-election forecasts made by polls were badly wrong [Bandera, (2016)]. Between them, the two main parties shed more than 5 million votes, while two new emerging parties erupted onto the political scene.

First, Podemos [Casero-Ripollés *et al.* (2016), Chadwick and Stromer-Galley, (2016)], which won more than one million votes and became the fourth largest party in Spain. Podemos stated that it was a party whose intention was to encompass the whole political spectrum and so, in the party's own words, to move on from the "old" dialectic between left and right. However, the truth is that what the party proposed only resulted in it being perceived as a wholly leftist party and for the European Council on Foreign Relations (ECFR, 2016), Podemos is an insurgent left-wing party. Second, a centre-right party, Ciudadanos (C's), which made the leap from Catalanian regional politics [see Bosch and Espasa, (2014) and Comerford *et al.* (2014), as contrasting examples of the current complexity of Catalanian Politics] to national politics with almost 500,000 votes. Ciudadanos originally defined itself as a social democratic party, although it has now (2017) changed its definition in the party statutes to liberal. Both new parties, Podemos and Ciudadanos, subsequently consolidated their positions in the 24 May 2015 regional and local elections [Rodon and Hierro, (2016)].

The following elections, held on 20 December 2015 (referred to as 20-D), produced an unprecedented outcome in recent Spanish democratic history: four political parties each received over 10 % of the valid votes cast, yet none obtained 30 %. The parties were unable to form a government in this multiparty scenario, forcing a rerun of the elections six months later, on 26 June 2016 (26-J), and dealing another blow to the already turbulent national political situation [Castillo-Manzano *et al.* (2017)].

The repeat of the elections, distrust in political parties –due as much to their not keeping their electoral promises as to corruption scandals– new emerging/insurgent parties and new candidates (three of the four main candidates for the post of prime minister were running for the first time in the 20-D elections) produced new circumstances and increased the electorate's volatility, with polls of voting intention presenting constantly changing scenarios.

On both 20-D and 26-J, once the polling stations had closed both the surveys conducted during the election campaign and the exit polls were seen to have been systematically wrong, especially in their predictions for the emerging parties. In the first election campaign (20-D), the polls indicated that the PP would get the most votes but did not go so far as to predict whether it would get a large enough majority to form a government, and second and third places were highly contested in the polls (Orriols and Cordero, 2016). To be more specific, the 20-D election results showed that the surveys had generally underestimated the number of votes that would be cast for the PP and Podemos, while greatly overestimating those that would be cast for Ciudadanos.

The 26-J polls once again badly miscalculated by underestimating the winning party, the PP, while on this occasion giving greater voting intention to an emerging party, the alliance between Podemos ("We Can") and the old Izquierda Unida ("the United Left"). Running together as Unidos Podemos ("United We Can"), the alliance

ended up in third place behind the PP and the PSOE (Socialist Party), preventing what had been termed a *sorpasso* (Italian, meaning “outflanking” or “overtaking”) in the political press.

These systematic errors in the Spanish polls, especially after the onset of the great recession, are not an isolated case in geographic terms. The US is a good illustration. In the US, the pre-election polls have worked quite well in the past; for example, the Bush victory and his winning margin over Kerry [Pickup and Johnston, (2008), Traugott, (2005)] were correctly predicted in 2004. The same was true of the 2008 presidential election, with the Obama victory over his Republican opponent McCain [Panagopoulos (2009)]. However, there has been an increasing number of sizable prediction errors in recent years; for example, in 2012, support for the Republican Romney was overestimated [Panagopoulos and Farrer, (2014)] and, more recently, in 2016, the polls were biased toward the Democratic candidate, Clinton [Anuta *et al.* (2017)].

Modern elections are characterized by the ubiquity of polls [Wei *et al.* (2011)], despite the criticism that they have received. In most countries there are many organizations that conduct polls, from newspapers to other media groups, not to mention think tanks, universities and dedicated market research companies [Panagopoulos and Farrer, (2014)]. However, these days pre-election and exit pollsters face major challenges, such as their procedures, including the estimation of likely voters, and technological advances that have paved the way to internet-based and interactive voice response polls. Such features require greater information to be provided about survey methods and the consequences that they might have on estimations, in order not only to maintain trust in the surveys themselves, but also in the transparency of the election process [Traugott, (2005)]. It is important for a society to believe in polls for, as Tremayne (2015) states, when part of the population sees that the polls into which they have put their trust have not been fulfilled, there can be accusations of voter fraud or political trickery and, ultimately, the foundations of the democratic system itself can be questioned.

Following Price and Stroud (2005), the credibility of polls lies in people believing that they are reasonably accurate soundings of popular opinion. There are several factors that impact the accuracy of polls and, therefore, their credibility. On the one hand, there are the technical characteristics and survey design. However, the difference between electoral predictions and results cannot simply be put down to a statistical error, but derive from sample design, question formulation, weighting, and screening [Pickup and Johnston, (2008)]. Second, there is the behavior of the respondents themselves. Not everybody is willing to state their voting intention to pollsters, whether due to reasons of confidentiality or privacy [Kim *et al.* (2011)] or simply because they are undecided or do not want to show which way they are going to vote. However, the votes of the undecided and the hidden vote are not always uniformly distributed among voters for the different parties and this could generate a problem of selection bias, which would also distort the survey [Urquizu-Sancho, (2006)]. In addition, respondents’ distrust of a survey is a process that feeds into itself. A negative perception or lack of faith in pre-election polls can lead citizens to lose their appetite for being poll respondents, which then affects the accuracy of the polls, and this would consequently increase the negative perception or lack of trust even more in the future.

Lastly, a large part of the distrust in the pre-election polls is caused by the belief that pollsters are trying to influence election outcomes [Price and Stroud, 2005]). By this we mean a bias that comes from the very polling organizations or companies themselves in surveys conducted or commissioned by media with a clearly defined editorial line [Eveland and Shah, (2003)]. Given that poll results are a core part of media coverage, the polls themselves often constitute electoral news in the media [Roy *et al.* (2015)], especially given today's demand for 24-hour news [Rosenstiel, (2005)]. As such, any analysis should include the effects that the political ideology of the company that commissions and/or conducts the polls might have on the selection of polls that are written about, the details that are emphasized, and the way that polling data are interpreted [Searles *et al.* (2016), Tremayne, (2015)].

For all these reasons, the main purpose of this paper is to utilize the Spanish case study to explain why polls are making greater errors at a time when society and its citizens want more from polls, i.e., in times of great economic slumps and political instability, and the emergence of new political parties. The paper will, therefore, focus on the factors that might determine their level of accuracy in the last two general elections (20-D and 26-J). In all the history of Spain's present-day democracy, the importance of voter expectation polls may well have peaked in the media during the above-mentioned time period. When significant changes in voter intention were announced, they quickly became front page news and the lead story in news bulletins. The Spanish people watched on as Podemos was declared the leading political party-in-waiting by voter intention one month, and Ciudadanos the next, with both outstripping the two great traditional parties. And yet, the two emerging parties in fact came in 3rd and 4th in both elections.

In short, we shall seek to use the Spanish case to respond to a broad series of questions such as, for example, whether polls are an efficient tool for reducing social uncertainty in precarious socioeconomic times when new political parties spring up. Or, for revealing the factors that would enable us to reduce errors in the polls during times of adversity like the current period. Other questions will also be addressed with managerial implications for the media, such as what the optimal economic strategy is in a context where the polls become front page news.

A earlier study of the Spanish case [Bandera, (2016)] examined the accuracy of the polls in the 2014 European elections, the autonomous regional elections of 2015 and the local elections of 2015, highlighting, as commented previously, that the first of these involved a change of cycle that none of the polling companies managed to capture in their surveys of the public. As Buchanan (1986) states, when the electorate is volatile, survey procedures based on experience in more stable situations may not work as well. It is therefore pertinent to analyze the accuracy of the pre-election surveys conducted for the general elections of 2015 and 2016 and to ascertain what the factors were that had influenced them, including, for the first time, the political bias of the company commissioning and/or conducting the survey.

1. DATA AND METHOD

As commented above, the purpose of this paper is to measure the source of the inaccuracy of the electoral polls, with inaccuracy of the polls understood as the difference between data prediction and reality.

The initial data come from the compilation made by the El Mundo newspaper (<http://www.elmundo.es/grafico/espana/2015/10/15/561fe19422601dd7728b45ef.html>), which collected data from 24 companies, organizations and media for 222 electoral polls between 6 January 2015 and 20 June 2016. Using this data we have looked up the original sources and corrected various mistakes in the El Mundo database, whilst also adding some new variables, such as the survey methods used. The database has also been expanded to include two exit polls and two tracking polls (one of each for the two elections) that were not published until the polling stations had closed, as Spanish law prohibits the publication of survey data during the last five days before ballots.

The El Mundo database was assumed to work with Unidos Podemos to standardize data in both the 20-D and 26-J elections, although this was a misrepresentation of the truth, as the coalition only existed during the second election. With the help of the original sources, we were able to recalculate all the mistakes made in the first election and so allow for the fact that Podemos ran alone in the first election.

The indicators taken to measure inaccuracy as dependent variables were M3 and M5, proposed by Panagopoulos and Farrer (2014) and based on Mosteller *et al.* (1949). Specifically, M3 is the average absolute difference between the poll estimate and the final election result for each party, while M5 compares the polled margin between the two main parties to the eventual outcome margin between the same parties and returns the absolute value of the difference between these margins. To enrich the analysis, in our case we have extrapolated the M5 concept not only to the main parties in the two elections, PP and PSOE, but have also calculated a specific M5 for the two emerging parties, Podemos/Unidos Podemos and Ciudadanos. A total of six different dependent variables will be worked with, four of which are related to measuring M3, and the other two to measuring M5.

The following independent variables in our model are factors that might influence survey errors. Their choice is justified by the prior literature [Crespi, (1988), Lau, (1994), Martin *et al.* (2005); Panagopoulos and Farrer, (2014)]:

a) *Sample Size* (Size). Whether the size of the sample might affect accuracy, as we a priori think it might. With this variable we will also implicitly be evaluating the need for the media that commission polls to bear a high cost for them to be conducted, as said cost should be closely correlated with sample size.

b) *The way that the poll is conducted* (Mode). Polls will be analyzed to determine whether they are based on face-to-face, telephone or online interviews. Current technological advances have enabled the growth of online surveys due to their lower-cost. In our case, we shall test whether the predominant mode, i.e., polls conducted by telephone interview, provide some sort of advantage over other polling methods.

c) *The lag between the survey and the day of the election* (Number_days). It can be assumed that the nearer the election date, the more certain respondents will be about the party that they plan to vote for and less inaccurate the survey will be. With this variable we are also testing the utility of polls at times when there are no elections in the near future.

d) One particularity of this case, based on two general elections in Spain within a period of only six months, is that we have considered it appropriate to include whether the polls were conducted for the *first or second election process* (Sec-

ond_time). A priori the polls' predictive ability should be expected to improve in the second election process, as the recent real results for the new political parties are already available for a political election of the same type.

e) We shall also test the *experience of the company* conducting the surveys through a variable that measures the number of polls that the company has conducted in the period under consideration (Number_polls). A priori the greater experience of the company should be expected to translate into a smaller prediction error in the polls.

f) As mentioned in the Introduction, the model also includes the possible *ideological bias* of the outlet that has commissioned the survey and, obviously, disseminated its results. A priori, this variable is especially relevant in the Spanish case [Van Dalen, (2012), Vliegenthart and Mena Montes, (2014)], as the media, and especially television, have played a major role in the rise of the new parties, Podemos and Ciudadanos. In fact, for the first elections in which Podemos ran, the European elections of 2014, instead of putting the party logo on the ballot papers, as all political parties traditionally do, they used a photo of their leader, Pablo Iglesias, to draw on his media appeal, as he frequently appeared on television talk shows and other TV programs.

Media bias was obtained through an online survey campaign of a panel of Spanish university experts in social sciences, all of whom have numerous publications in Scopus, especially in the political and economic sciences. Specifically, they were asked whether the 24 different media outlets included in our analysis have a political editorial bias toward supporting right-wing parties, left-wing parties, or whether it was neutral, whilst also offering the option of "do not know/no answer". We received a total of 36 responses, four of which were ruled out as they lacked data, leaving 32 valid results. The researchers in question stated that they originated from 11 of the 17 Spanish autonomous communities (regions), although 4 preferred not to answer this question.

A very strict definition was used to define biases. An outlet was considered to be biased toward one ideological wing or the other when responses indicating this exceeded the sum of the responses indicating the opposite bias and any responses indicating a neutral mid-ground. Based on this definition, only 15 of the 24 analyzed outlets presented a clear ideological bias in their editorial line; 10 to the right, and 5 to the left. These results reveal that a set of media outlets exists in the Spanish political arena with a definite right-wing bias.

The variables Left_Bias and Right_Bias were calculated on the basis of the results of this process. The first variable, Left_Bias, is a dummy variable that takes a value of 1 if the outlet that commissioned and disseminated the survey has a clear left-wing ideological bias according to the expert panel; otherwise, it takes a value of 0 if there is no clear political bias. The Right_Bias variable is defined in the same way.

It is not easy to determine a priori the direction of the influence of these two biases, Left and Right. It might be thought that outlets with a specific bias could attempt to give some advantage to the results of parties close to their political way of thinking in order to get inside voters who like to back the winner, or to underestimate the results in order to mobilize a sympathetic, but possibly apathetic, electorate. A strategy like this is based on the theory that voters cast their votes not only on the basis of their preferences, but also their perceptions of the parties' chances of winning [Blais and Bodet, (2006)].

Another important aspect is the role that Right Bias media have played with regard to an emerging left-wing insurgent party like Podemos. Furthermore, there have been frequent clashes and spats between the leaders of Podemos and some of the journalists from these media. Therefore, it comes as no surprise that Right Bias media have repeatedly defined Podemos as an anti-system populist party and have flagged up some possible links between the party's leaders and countries like Iran and, especially, Venezuela. In spite of this, Podemos' leaders have overcome all the allegations leveled against them and investigations into the party allegedly receiving illegal funding from abroad.

g) *Sampling_error*. A dummy variable which takes a value of 1 if electoral poll i provides information about its sampling error and 0 otherwise. This variable is an indicator not only of the quality and rigor with which the poll has been conducted but also of its transparency. This also enables us to test the Mateos and Penadés (2013) hypothesis that one of the key factors that explains the reliability of polls in predicting the vote is not just the size of the poll but also its transparency.

h) *Stock_Market*. Also included as a dummy variable, taking a value of 1 if electoral poll i was commissioned by a media outlet or a news website with a parent firm listed on the stock market, and 0 otherwise. This variable was included to enable us to discriminate between the media that can a priori count on greater economic resources, which would enable them to conduct rigorous polls.

i) Lastly, we have included a variable that seeks to quantify the influence of the *Environment*, i.e., the influence that the results of contemporaneous polls have.

Table 1 provides a description of both the endogenous and explanatory variables used, together with their descriptive statistics.

2. RESULTS AND DISCUSSION

Table 2 presents the main results of the estimates. Specifically, each column is a linear regression for a specific endogenous variable which is made explicit in the column.

To test for any possible multicollinearity problems that might exaggerate estimates of the variance parameter and distort its statistical significance, we have calculated the Variance Inflation Factors (VIF) of the variables used in the empirical analysis. The usual rule-of-thumb considered in econometric textbooks is 10, although practitioners may use lower threshold values and a value of 5 is common. In our case, the values of the VIFs are very low. Specifically, the average is 1.47 with all values below 3. In fact, they are practically all below 2.

It is easy to observe (see Table 1) that when there are flourishing emerging parties such as Podemos and Ciudadanos, election poll errors for these parties increase exponentially, as can be seen in a comparison of, e.g., the value of the mean of $M3_New_i$ and the $M3_Trad_i$ variables, specifically 3.676% versus 2.324%. In some models, especially No. VI, the constant has a quite persistent error that the polls have to contend with in times of great uncertainty, as described in this paper. It was only possible for this error to be moderated in the second election (*Second_time* variable).

In addition, and even more importantly, no optimal solution seems to exist that would enable the polling companies to reduce these errors. Variables *Size* and *Num-*

Table 1: DESCRIPTION OF THE VARIABLES AND THEIR DESCRIPTIVE STATISTICS

Variable	No. Obs. (only dummies = 1)	Mean	Median	Std. dev.
1. Endogenous variable				
<i>M3_Trad_i</i> : Mean of absolute errors made by survey <i>i</i> in the prediction of the election results of the two main parties (PP and PSOE).	–	2.324	2.22	1.112
<i>M3_New_i</i> : Mean of absolute errors made by survey <i>i</i> in the prediction of the election results of the two main emerging parties (Podemos/Unidos Podemos and Ciudadanos).	–	3.676	3.3	1.719
<i>M3_Left_i</i> : Mean of absolute errors made by survey <i>i</i> in the prediction of the election results of the two main left-wing parties (PSOE and Podemos/Unidos Podemos).	–	2.803	2.63	1.449
<i>M3_Right_i</i> : Mean of absolute errors made by survey <i>i</i> in the prediction of the election results of the two main liberal or right-wing parties (PP and Ciudadanos).	–	3.200	2.975	1.542
<i>M5_Trad_i</i> : Absolute value of the difference between polled margins of survey <i>i</i> and outcome margins of the two traditional main parties (PP and PSOE), which also received the most votes in both elections.	–	2.632	2.31	1.950
<i>M5_New_i</i> : Absolute value of the difference between polled margins of survey <i>i</i> and outcome margins of the two main emerging parties (Podemos/Unidos Podemos and Ciudadanos), which were in 3rd and 4th places.	–	5.268	3.72	4.420
2. Explanatory variables				
<i>Size_i</i> : Size of survey <i>i</i> , specifically, the number of people interviewed (in thousands) in said survey.	–	3.031	1.1	14.758
<i>Mode_i</i> : Dummy variable, which takes a value of 1 if electoral poll <i>i</i> was conducted by telephone interview and 0 otherwise (face-to-face interview, over the internet, etc.).	198	0.892	1	0.311
<i>Number_days_i</i> : Number of days from the presentation of the results of survey <i>i</i> to the date on which the general election was held.	–	105.324	79	93.018

Table 1: DESCRIPTION OF THE VARIABLES AND THEIR DESCRIPTIVE STATISTICS (continuation)

Variable	No. Obs. (only dummies = 1)	Mean	Median	Std. dev.
2. Explanatory variables				
<i>Second_time_i</i> : Dummy variable which takes a value of 1 if survey <i>i</i> refers to the second general election (June 2016), and 0 if it refers to the first (December 2015).	87	0.392	0	0.489
<i>Number_polls_i</i> : Number of electoral polls conducted by the company carrying out electoral poll <i>i</i> during the period under study.	–	18.018	18	8.197
<i>Left_Bias_i</i> : Dummy variable which takes a value of 1 if the media outlet that commissioned and disseminated survey <i>i</i> has a clear left-wing ideological bias.	72	0.324	0	0.469
<i>Right_Bias_i</i> : Dummy variable which takes a value of 1 if the media outlet that commissioned and disseminated survey <i>i</i> has a clear right-wing ideological bias.	93	0.419	0	0.495
<i>Sampling_error_i</i> : Dummy variable which takes a value of 1 if electoral poll <i>i</i> provides information about its sampling error and 0 otherwise.	168	0.757	1	0.430
<i>Stock_Market_i</i> : Dummy variable which takes a value of 1 if electoral poll <i>i</i> was commissioned by a media outlet or website with a parent firm listed on the stock market and 0 otherwise.	110	0.495	0	0.501
<i>Environment_i</i> : Average value of endogenous variable in the 5 polls simultaneous with or previous to <i>i</i> . The calculation process was restarted after the 20D elections	Model I	2.359	2.377	0.652
	Model II	3.745	3.223	1.372
	Model III	2.840	2.642	0.830
	Model IV	3.261	3.035	1.050
	Model V	2.669	2.562	1.137
	Model VI	5.395	3.622	4.085

Table 2: ESTIMATES FOR THE DIFFERENT DEPENDENT VARIABLES (IN BRACKETS)

Independent variables	Model I (M3_Tradi)	Model II (M3_New _t)	Model III (M3_Left _t)	Model IV (M3_Right _t)	Model V (M5_Tradi)	Model VI (M5_New _t)
<i>Size_i</i>	-0.003 (0.002)	-0.010 (0.007)	-0.007 (0.004)	-0.009** (0.004)	-0.014** (0.006)	-0.016 (0.023)
<i>Mode_i</i>	-0.139 (0.219)	0.473 (0.324)	-0.173 (0.307)	0.262 (0.369)	-0.761 (0.389)	1.430** (0.715)
<i>Number_days_i</i>	0.003** (0.001)	-0.002 (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.001 (0.002)	-0.005* (0.003)
<i>Second_time_i</i>	0.586*** (0.175)	-0.675*** (0.245)	-0.676*** (0.237)	0.052 (0.163)	0.363 (0.247)	-2.259*** (0.606)
<i>Number_polls_i</i>	-0.011 (0.008)	0.001 (0.012)	0.010 (0.011)	-0.018 (0.011)	0.004 (0.014)	-0.016 (0.025)
<i>Left_Bias_i</i>	-0.272 (0.183)	0.238 (0.250)	0.400 (0.247)	-0.482* (0.263)	-1.140*** (0.382)	0.407 (0.535)
<i>Right_Bias_i</i>	-0.801*** (0.173)	-0.486** (0.235)	-0.032 (0.213)	-1.258*** (0.250)	-2.054*** (0.360)	-0.435 (0.491)
<i>Sampling_Error_i</i>	-0.029 (0.128)	-0.039 (0.197)	-0.109 (0.226)	-0.044 (0.197)	0.298 (0.273)	-0.317 (0.433)
<i>Stock_Market_i</i>	0.079 (0.125)	0.094 (0.190)	0.181 (0.184)	0.045 (0.165)	0.421** (0.209)	-0.040 (0.390)
<i>Environment_i</i>	0.333** (0.159)	0.555*** (0.097)	0.224 (0.185)	0.688*** (0.094)	0.573*** (0.159)	0.624*** (0.074)
<i>Constant</i>	1.797*** (0.349)	1.737*** (0.533)	2.347*** (0.775)	1.860*** (0.435)	2.320*** (0.474)	2.620*** (0.984)
R ²	0.362	0.381	0.141	0.392	0.368	0.598
Test joint significance	12.52***	17.60***	4.13***	15.43***	14.49***	47.40***
Number of observations	218	215	215	218	218	215
Root MSE	0.9095	1.359	1.376	1.212	1.593	2.8295

Note 1: Standard errors robust to heteroscedasticity in brackets. Statistical significance at 1% (***), 5% (**), 10% (*).

ber_polls are not significant for the emerging parties in models II and VI. This shows that any survey is bound to be inaccurate when new options exist that command major social support, irrespective of how many times the survey is repeated or how broad the sample is.

These strategies are not very effective for the traditional parties, either. Although the variable *Size* generally presents a negative sign as was expected (the greater the survey sample size, the lower the error in the predictions), the only significant effects worth highlighting, at 5%, were in the measurement of traditional parties' M5 (Model V) and also in *M3_Right_i* (Model IV). The irrelevance of the sample size for predicting election results was also noted by Bandera (2016), Crespi (1988), Lau (1994), and Panagopoulos and Farrer (2014). In short, there seems to be no clear incentive for the media to undertake costly electoral polls with thousands of surveys. The strategy followed by some media, of sharing the cost of polls in order to increase the sample size, would also seem to be flawed. Only if said media are not competing for the same client (such as in the case of regional television stations), is it worth their while to commission a joint poll. Otherwise, the optimal strategy for a media outlet is to commission its own poll with few observations.

Meanwhile, the variable *Number_polls* is not significant for the traditional parties, either, meaning that, to the contrary of what might have been anticipated, experience does not count for the parties that are already well-known. In other words, if the methodology that a specific opinion poll company uses fails once, irrespective of the party in question, it will fail systematically.

The same is true for the variables *Sampling_Error_i* and *Stock_Market_i*. Their general lack of significance for both traditional and new parties strongly suggests that neither greater technical quality, nor transparency, nor the availability a priori of greater economic resources guarantee better predictions of the results.

Nonetheless, it would be a good strategy for traditional parties to conduct polls nearer the election date, as the variable *Number_days* is relevant for M3 in model I. As could be anticipated, surveys carried out as near to the election day as possible were those that were most correct. This result is in line with Crespi (1988), who found that the accuracy of polls depended on scheduling them as close as possible to the date of the election. However, it is also true that other studies find no significant impact of the timing of the poll on accuracy [Bandera, (2016), Lau, (1994), Panagopoulos, (2009), Panagopoulos and Farrer, (2014)]

However, this strategy does not work for emerging parties (Model II and Model VI), as prediction error is insensitive to the nearness of the date of the election. This result once again highlights the difficulty that the polls have in predicting voters' real behavior in the face of these new parties. Not even the closeness of the election, i.e., when in theory the voters should have a clearer idea about which party they are going to vote for, helps to reduce uncertainty.

For new parties, the key lies in the variable *Second_time* which is significant in four of the six models. Focusing on the emerging parties, it could be said that the only solution for reducing prediction errors for these parties is to hold some elections in advance and close in time, which is a highly unlikely scenario, as the Spanish case—with the election being repeated in under a year—is exceptional—. In short, new parties have to stop being “new” for their prediction error to approach some-

thing similar to that of traditional parties. Obviously, this means that polling companies can use data on past voting behavior to correct sample biases and as a starting point in their forecasting models.

In our case, the estimates tell us that the polls for the second election achieved an immediate reduction of 0.675 percentage points in the mean error (M3) for the new parties, whilst at the same time there was a 0.586 increase for the traditional parties. As far as M5 is concerned, the error reduces even more, by -2.259 percentage points. In short, when there are emerging parties in the mix, getting the real voting intention right is a complicated matter, and it does not appear that polls are reducing this logical uncertainty. Only real elections help clarify the landscape.

In this regard, certain parallels exist between the great errors of Spanish polls and those of Brexit and the election of Donald Trump [Chadwick and Stromer-Galley, (2016)]. In all three cases voters had to choose between brand new choices with no recent precedent [Bachmann and Sidaway, (2016) on Brexit; in the case of the US election, although he was running for the Republican Party, Trump was obviously not a model of a conventional politician]. In other words, in scenarios with parties and/or emerging political options, societies are condemned to swallow large doses of uncertainty for which the polls offer no efficient solutions.

With respect to the other explanatory variables used, it can be concluded that, first, there has been no greater accuracy shown by the use of telephone interviews over other types of survey, such as face-to-face or online interviews, except in the case of the M5 prediction for traditional parties (see variable *Mode*). This is in line with what was observed in Panagopoulos and Farrer (2014) but unlike the results found in Panagopoulos (2009), which showed that polls conducted via the Internet were significantly less accurate than polls conducted by telephone. In short, this result supports the hypothesis that polling companies use whichever mode is cheapest for them as long as it guarantees that they obtain a real simple random sample.

Second, there seems to be a clear bias toward the average of preceding polls. In fact, in practically all the estimated models, *Environment_t* is usually the main explanatory variable. This result would indicate that pollsters present a certain conservative behavior in their predictions, bringing them more into line with what their competitors present at the same moment in time.

Finally, editorial bias (*Left_Bias* and *Right_Bias*) appears to play a significant role, especially in the case of traditional parties. It was the media with a clearly politicized editorial line, especially those on the right-wing, that made the best M3 and M5 predictions, especially for the traditional parties and for the parties on the right. This result justifies the inclusion of this type of variable in the analysis as it would ultimately be capturing the influence that the media have in individual countries.

It should be noted that all the correlations, whenever they exist (Models I, II, IV and V), are always negative. That is to say, the political bias of the outlet reduces the polling error in the case of both *Left_Bias* and *Right_Bias*. This result could serve as indirect confirmation that the media apparently do not deliberately or systematically implement any clearly biased practices to tamper with the electoral poll results in order to favor parties with the same ideology.

Whatever the case, a future line of research would be to find explanations for this minor prediction error of biased media. One possible hypothesis might be that

the variables *Left_Bias* and *Right_Bias* are really capturing some specific characteristics of the media other than their ideological bias in the Spanish case. It is noticeable that many of the most eminent media, with the most experience and, above all, with the most human and economic resources (such as ABC, Antena 3, El Mundo, La Vanguardia and SER) are among those that have been identified as ideologically biased by the panel of experts that we have employed.

3. CONCLUSION

Leaving aside the initial instability from the late 1960's until the failed coup d'état in 1981, the Great Recession that began in 2008 has been the main challenge that Spanish democracy has had to face. In this context of economic, social and political crisis, polls have become immensely important, as they are, theoretically, the instrument that should attempt to mitigate the political uncertainty that quickly spread with the emergence of new political parties trying to challenge the traditional two-party system.

However, the polls offered significantly different results month on month, even changing the most voted party, and even offering implausible technical links among the four main parties, right up to a month before the election. To this must be added their lack of success in predicting the final election results. For all these reasons, their real utility was called into question even further, and this did not exactly improve trust in the Spanish political system.

Following Fisher *et al.* (2011), the proliferation of electoral poll data should make us rigorously assess what we can learn about the results of an election from voter intention polls. The present paper has therefore sought to explain their lack of success in the Spanish case through an econometric analysis of 226 electoral polls, and has attempted to define some recommendations that would enable an optimal strategy to be determined for such polls when the situation is one of absolute uncertainty.

Unfortunately, the findings have shown that reducing the electoral polls' margin of error is a difficult proposition in such a volatile context. The prediction error for the new emerging parties is significantly higher than for the traditional parties. And what is even worse, this error in the estimation of the new parties is not sensitive to solutions for increasing the reliability of surveys, such as increasing sample size, transparency or constantly conducting periodical surveys. Not even the closeness of the approaching election significantly reduces this error. Only when the emerging parties are no longer "new" will it be possible to reduce this error, i.e., when they have taken part in similar elections (and, it appears, different types of elections are not enough, since in 2015 Podemos had taken part in two elections only a few months before the general election of 20-D: European and municipal-regional elections, to be precise). Be that as it may, it is only fair to recognize the technical difficulty involved in a poll being able to predict the emergence of new social phenomena (such as new parties). However, when they are out there, polls nevertheless seem to do a good job of learning how to predict their vote share, even if they still make better predictions for traditional parties.

On the other hand, the determinants of the reliability of the predictions for traditional parties do follow a pattern that is more consistent with what is stated in the

prior literature. To begin with, the predictions in this case are, on average, a little more reliable and sensitive to sample size and to the time remaining until the election.

Other general conclusions of this study are, first, the utility of including the medium's editorial bias, even though, once more, it influences mostly predictions for traditional parties. Second, the apparent lack of utility of the strategy that many media follow of systematically making predictions every so often. In other words, this greater experience (and economic cost) does not translate into greater trustworthiness, although it might generate more original headlines that are not available to the competition, as only the outlet that commissions the poll can exclusively disseminate the results.

In fact, this is a phenomenon that feeds into itself and creates a vicious circle. The political uncertainty derived from the crisis and the new parties leads citizens to crave more electoral polls. The media see a good business opportunity in this need. They have many polls carried out, with differing results that change in the short term and which are not very accurate compared to the real results, which contributes to generating greater uncertainty and a greater lack of trust in the system. And this, in turn, generates further social uncertainty, which leads to a greater demand for polls.

In this context, it is the very instrument with which society tries to reduce the reigning uncertainty that, paradoxically, can ultimately drive uncertainty up. Therefore, the best recommendation is that the media that present the polls should stress their limitations when new parties and a severe socio-economic crisis are involved. This would prevent unfounded expectations of their reliability and, above all, shocks to society for no apparent real reason.

The best service that opinion poll companies and the media that contract them can offer society is to modestly acknowledge the limitations of their electoral polls.

It can be observed that pollsters also seem to present clear risk aversion behavior, in the sense that they do not want to make predictions that vary greatly from the average of the other polls. This strategy must be defensive in part as, even if the average is wrong, reputational blame can be avoided if all the other pollsters also fail.

To conclude, these results also have implications for epistemology in the social sciences. On occasion the results of the polls are used as a dependent variable in models [Guiso *et al.* (2017), for example, where they are used to measure the increase in populism in Europe] under the assumption that they reliably represent society's political preferences at any given point in time. It would be desirable for these models to include some kind of correction that would take into account the fact that the reliability of results diminishes with the emergence of new vibrant political parties [Welch, (2002) on disclosing essential information about polls].



REFERENCES

- Anuta, D., Churchin, J. and Luo, J. (2017): "Election Bias: Comparing Polls and Twitter in the 2016 U.S. Election", Cornell University Library Working Paper (January).
- Bachmann, V. and Sidaway, J.D. (2016): "Brexit geopolitics". *Geoforum*, vol. 77, pp. 47-50.
- Bandera, N. (2016): "How accurate have the polls been in the new political period? 2014 European elections and 2015 regional/local Spanish elections *EMPIRIA*", *Revista de Metodología de Ciencias Sociales*, n. 34, pp. 124-154.

- Blais, A. and Bodet, M.A. (2006): "How Do Voters Form Expectations about the Parties' Chances of Winning the Election?" *Social Science Quarterly*, vol. 87, pp. 477-493.
- Bosh, N. and Espasa, M. (2014): "La viabilidad económica de una Cataluña independiente", *Revista de Economía Aplicada*, n. 64, pp. 135-162.
- Buchanan, W. (1986): "Election Predictions: An Empirical Assessment", *The Public Opinion Quarterly*, vol 50, pp. 222-227.
- Camacho M., Gadea M.D. and Pérez-Quirós, G. (2018): "The Great Recession: the worst ever?", *Revista de Economía Aplicada* n. 76, pp. 73-100.
- Casero-Ripollés, A., Feenstra, R.A. and Tormey, S. (2016): "Old and New Media Logics in an Electoral Campaign: The Case of Podemos and the Two-Way Street Mediatization of Politics" *The International Journal of Press/Politics*, vol. 21, pp. 378-397.
- Castillo-Manzano, J.I., Lopez-Valpuesta, L. and Pozo-Barajas, R. (2017): "Six months and two parliamentary elections in Spain: December, 2015 and June, 2016. The end of the two-party system?" *Electoral Studies* vol. 45, pp. 157-160.
- Chadwick, A. and Stromer-Galley, J. (2016): "Digital Media, Power, and Democracy in Parties and Election Campaigns: Party Decline or Party Renewal?" *The International Journal of Press/Politics*, vol. 21, pp. 283-293.
- Chari, R. (2013): "The parliamentary election in Spain, November 2011", *Electoral Studies*, vol. 32, pp. 378-381.
- Clarke, N. (2015): "Geographies of politics and anti-politics", *Geoforum* vol.62, pp. 190-192.
- Crespi, I. (1988): *Pre-Election Polling: Sources of Accuracy and Error*, New York, Russell Sage Foundation.
- Comerford, D., Myers, N. and Rodríguez Mora, J.V. (2014): "Aspectos comerciales y fiscales relevantes para evaluar las consecuencias económicas de una hipotética independencia de Cataluña", *Revista de Economía Aplicada*, n. 64, pp. 85-130.
- ECFR. (2016): "The world according to Europe's insurgent parties: Putin, migration and people power", *European Council on Foreign Relations* (June).
- Eveland, W.P. and Shah, D.V. (2003). "The Impact of Individual and Interpersonal Factors on Perceived News Media Bias", *Political Psychology*, vol. 24, pp. 101-117.
- Fisher, S.D., Ford, R., Jennings, W., Pickup, M. and Wlezien, C. (2011). "From polls to votes to seats: Forecasting the 2010 British general election", *Electoral Studies*, vol. 30, pp. 250-257.
- García Serrano, C. (2011): "Déjà vu? Crisis de empleo y reformas laborales en España", *Revista de Economía Aplicada*, n. 56, vol. 149-177.
- Guiso, L., Herrera, H., Morelli, M. and Sonno, T. (2017): "Demand and Supply of Populism. Centre for Economic Policy Research (CEPR)", Working paper DP11871 (February).
- Hameleers, M., Bos, L. and de Vreese, C.H. (2017): "The Appeal of Media Populism: The Media Preferences of Citizens with Populist Attitudes", *Mass Communication and Society*, vol. 20, pp. 481-504.
- Kim, J., Gershenson, C., Glaser, P. and Smith T.W. (2011): "The Polls-trends trends in surveys on surveys", *Public Opinion Quarterly*, vol. 75, pp. 165-191.
- Lau, R.R. (1994): "An analysis of the accuracy of "trial heat" polls during the 1992 presidential election", *Public Opinion Quarterly*, vol. 58, pp. 2-20.
- Martin, E., Traugott, M. and Kennedy, C. (2005): "A Review and Proposal for a New Measure of Poll Accuracy", *Public Opinion Quarterly*, vol. 69, pp. 342-369.
- Mateos, A. and Penadés, A. (2013): Las encuestas electorales en la prensa escrita (2008-2011). Errores, sesgos y transparencia, *Metodología de Encuestas*, vol. 15, pp. 99-119.
- Mosteller, F., Hyman, H., McCarthy, P., Marks, E. and Truman, D. (1949): *The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-Election Polls and Forecasts*. New York: Social Science Research Council.

- Orriols, L. and Cordero, G. (2016): "The Breakdown of the Spanish Two Party System: The Upsurge of Podemos and Ciudadanos in the 2015 General Election", *South European Society and Politics*, vol. 21, pp. 469-492.
- Panagopoulos, C. (2009). "Polls and elections: preelection poll accuracy in the 2008 general elections", *Presidential Studies Quarterly*, vol. 39, pp. 896-907.
- Panagopoulos, C. and Farrer, B. (2014). "Preelection Poll Accuracy and Bias in the 2012 General Elections", *Presidential Studies Quarterly*, vol. 44, pp. 352-363.
- Pickup, M. and Johnston, R. (2008): "Campaign trial heats as election forecasts: Measurement error and bias in 2004 presidential campaign polls", *International Journal of Forecasting*, vol. 24, pp. 272-284.
- Price, V. and Stroud, N.J. (2005): "Public attitudes toward polls: evidence from the 2000 U.S. Presidential Election", *International Journal of Public Opinion Research*, vol. 18, pp. 393-421.
- Rodon, T. and Hierro, M.J. (2016): "Podemos and Ciudadanos Shake up the Spanish Party System: The 2015 Local and Regional Elections", *South European Society and Politics*, vol. 2, pp. 339-357.
- Rosenstiel, T. (2005). "Political polling and the new media culture: a case of more being less", *Public Opinion Quarterly (2005 Special Issue)*, vol. 69, pp. 698-715.
- Roy, J., Singh, S.P., Fournier, P., and Andrew, B. (2015): "An experimental analysis of the impact of campaign polls on electoral information seeking", *Electoral Studies*, pp. 146-157.
- Searles, K., Ginn, M.H. and Nickens, J. (2016): "For whom the poll airs: comparing poll results to television poll coverage", *Public Opinion Quarterly*, vol. 80, pp. 943-963.
- Traugott, M. (2005): "The Accuracy of the National Preelection Polls in the 2004 Presidential Election", *Public Opinion Quarterly*, vol. 69, pp. 642-54.
- Tremayne, M. (2015): "Partisan Media and Political Poll Coverage", *Journal of Information Technology & Politics*, vol. 12, pp. 270-284.
- Urquizu-Sancho, I. (2006): "The non-declared vote in the surveys: The Spanish case in the 1980s", *Electoral Studies*, vol. 25, pp. 103-128.
- Van Dalen, A. (2012): "Structural Bias in Cross-National Perspective: How Political Systems and Journalism Cultures Influence Government Dominance in the News". *The International Journal of Press/Politics*, vol.17, pp. 32-55.
- Vliegenthart, R., and Mena Montes, N. (2014): "How Political and Media System Characteristics Moderate Interactions between Newspapers and Parliaments: Economic Crisis Attention in Spain and the Netherlands", *The International Journal of Press/Politics*, vol. 19, pp. 318-339.
- Wei, R., Chia, S.C. and V-H. Lo. (2011): "Third-person Effect and Hostile Media Perception Influences on Voter Attitudes toward Polls in the 2008 U.S. Presidential Election", *International Journal of Public Opinion Research*, vol. 23, pp. 169-190.
- Welch, R.L. (2002). "Polls, Polls, and More Polls. An Evaluation of How Public Opinion Polls Are Reported in Newspapers", *The International Journal of Press/Politics*, vol. 7, pp. 102-114.

Fecha de recepción del original: septiembre, 2018

Versión final: noviembre, 2018

RESUMEN

El principal objetivo de este artículo es explicar por qué las encuestas electorales cometen más errores en épocas de crisis económica, inestabilidad política y con partidos emergentes como Podemos y Ciudadanos. Para ello utilizamos una base de datos de 226 encuestas previas a las elecciones generales españolas de 2016. En este contexto, paradójicamente vemos como el instrumento que la sociedad utiliza para reducir su incertidumbre puede acabar aumentándola. Nuestros resultados muestran como el error de predicción de los nuevos partidos es significativamente mayor que los tradicionales e insensible a las soluciones clásicas para aumentar la precisión de las encuestas, como el tamaño de la muestra, el método de muestreo, la experiencia del encuestador, o la proximidad de la cita electoral. Además, se observa que las empresas que desarrollan las encuestas realizan de forma sistemática predicciones muy próximas a las que han realizado las encuestas recientes de sus competidores. Finalmente, el sesgo editorial parece ser una variable relevante, especialmente en lo relativo a las predicciones de los partidos tradicionales.

Palabras clave: encuestas electorales, partidos insurgentes, error de la encuesta, España.

Clasificación JEL: D7 - Análisis de la toma de decisiones colectiva.

