doi:10.5477/cis/reis.137.237

Nonresponse Bias and Superpopulation Models in Electoral Polls

Sesgo de no-respuesta y modelos de superpoblación en encuestas electorales

José M. Pavía y Beatriz Larraz

Key words

Electoral forecasts • Voting behaviour • Recall vote • Cluster Sampling • Pre-election polls • Monte Carlo simulation • Spanish elections

Abstract

Nonresponse bias (and, to a lesser extent, measurement error) has become the main source of error for electoral forecasts in Spain. Although the post-stratification techniques and ratio estimators currently used in the polling industry reduce deviations, they do not show enough capacity to mend the biases introduced when collecting data. This research reveals how a more efficient use of the electoral information available outside the sample could help to significantly improve the accuracy of predictions, and uses simulation techniques to show that this may be accompanied by less expensive sampling designs. The analysis, nevertheless, also concludes that the proposed specification is not a panacea and affirms that there is still scope for reducing nonresponse bias, pointing to several issues for future research.

Palabras clave

Predicciones electorales • Comportamiento electoral • Recuerdo de voto • Muestreo por conglomerados • Encuestas preelectorales • Simulación de Monte Carlo • Elecciones españolas

Resumen

El sesgo de no-respuesta (y, en menor medida, el error de respuesta) se ha convertido en la principal fuente de error de las predicciones electorales en España. Las técnicas de post-estratificación y los estimadores ratio utilizados actualmente por la industria demoscópica no muestran una capacidad suficiente para corregir los sesgos introducidos durante la recogida de datos. Este trabajo revela cómo un uso más eficiente de la información electoral extramuestral disponible permitiría mejorar sensiblemente la precisión de las estimaciones y muestra, utilizando técnicas de simulación, que ello podría venir acompañado de diseños muestrales más baratos. El estudio, no obstante, concluye que la especificación utilizada en esta investigación no constituye una panacea y señala que existe todavía margen para la corrección del sesgo de no-respuesta, apuntando diversas posibilidades de investigación futura.

Introducción[1]

Concern about the potential consequences of nonresponse[2] in survey research is as old as the discipline itself. According to Smith(1999), "*Early research extends back to the emergence of polling in the 1930s and has been a regular feature in statistical and social*

[1] The research reported in this paper is based on the study "Un Análisis para la Mejora de la Calidad Predictiva en las Estimaciones Electorales a partir de Datos de Encuesta" funded by the Centro de Investigaciones Sociológicas through a Sociological Research Grant awarded in 2009. The authors wish to thank Alberto Penadés and Valentín Martínez for their first-rate assistance, two anonymous referees and the REIS Editorial Board for their valuable comments and suggestions, and Tony Little for revising the English of the manuscript. Any

José M. Pavía: Universitat de València | Jose.M.Pavia@uv.es
Beatriz Larraz: Universidad de Castilla-La Mancha | Beariz.Larraz@uclm.es

science journals since the 1940s." Despite this, it has not been until more recently that the scientific community has devoted more attention to the issue, in response to the consequences that the increasing lack of willingness to participate in polls shown by many citizens is having on the quality of survey outcomes (Singer, 2006).

Since the second half of the 1980s, there has been a constant and gradual increase in surveying nonresponse rates, non-cooperation of respondents being established as the main cause of this trend (de Leeuw and de Heer, 2002). The efforts of researchers were promptly focused on trying to understand the causes of the phenomenon and on introducing new ideas, like the concept of randomness of nonresponse, in order to try to reduce nonresponse (in its two forms: non-contacted and rejections) and/or to correct its consequences (e.g., Groves and Couper, 1998). Social scientists and survey leaders directed their attention to understanding and reducing nonresponse (e.g., Groves *et al.*, 1999), while from a statistical perspective, research focused on trying to minimize its impact using techniques such as multiple imputation (e.g., Schafer, 1997; King *et al.*, 2001) and adjustment and weighting methods (e.g., Isaki *et al.*, 2004).

Nevertheless, once i) the inefficiency of the costly methods used to attempt to increase participation was recognized (e.g., Curtin *et al.*, 2005), ii) it was accepted, in the English literature (e.g., Keeter *et al.*, 2000; Merkle and Edelman, 2002; Groves, 2006), the lack of a stable relationship between the nonresponse rate and estimation bias and iii) it was acknowledged that the statistical methods currently in use are not capable of correcting sufficiently nonresponse bias[3], the main challenges that, according to Groves *et al.* (2002), survey methodology faced at the beginning of this century were to i) determine under what circumstances nonresponse can damage population inferences and ii) identify the methods that, in the presence of nonresponse, can improve the quality of sampling estimates. This paper aims to provide some answers, within the context of election forecasting, to the second of the challenges identified in Groves *et al.* (2002).

Within surveys, election polls and election predictions play an important role as, unlike most surveys, they may be judged against an external standard of comparison: the actual election outcomes. The sociological and public opinion forming aspects of election forecasts therefore help to shape the image of the whole sector (Martin *et al.*, 2005). Despite this, an analysis of the methods currently employed to generate election predictions reveals that the electoral information available outside the sample is used inefficiently. Indeed, the data recorded in previous elections could be more intensely exploited using superpolulation models (e.g., Valliant *et al.*, 2000).[4]

remaining errors are the authors sole responsibility. The authors acknowledge the support of the Spanish Ministry of Science and Innovation (MICINN) through the project CSO2009-11246/CPOL.

[2] In general, two types of nonresponse can be found in surveys: total or partial. Total nonresponse (or non-participation) occurs either when an individual cannot be reached to be interviewed (whose distribution among the various political options is often assumed as basically random) or when, after being contacted, s/he declined to be interviewed. This second type of nonresponse often shows a skewed distribution among the various options and is the main source of "*nonresponse bias*". Partial nonresponse (or item nonresponse) appears when the subject agrees to be interviewed but provided no answer to certain questions.

[3] This was painfully clear in the exit-poll predictions made in the 2000 U.S. Presidential Election, when errors, mainly attributed to nonresponse bias (e.g., Konner, 2003; Biemer *et al.*, 2003; Randall, 2008), provoked a great stir.

[4] In a superpopulation framework, the target population is considered a realization of a larger underlying population (superpopulation), where the individual realizations of each member of the population show certain patterns of regularity that can be statistically exploited and, therefore, be used to improve the quality of forecasts. From

In this line of research, when working with biased samples of counted votes, the forecasting election models based on the congruence that electoral outcomes of consecutive elections display at polling station or electoral section levels have shown great capacity to improve the accuracy of predictions (e.g., Bernardo and Girón, 1992; Bernardo, 1997; Pavía-Miralles, 2005; Pavía *et al*., 2008) even if only partial results from the polled stations are available (Pavía-Miralles and Larraz-Iribas, 2008). And, furthermore, this approach has very recently revealed its potential with data from exit polls (Pavía, 2010). The aim of this paper is to study and analyze the prediction capacity of this strategy when working with election polling data and to compare it to the procedures currently in use in the polling industry, more specifically, to post-stratification methods and ratio estimators (e.g., Mitofsky and Murray, 2002; Mitofsky, 2003).

In particular, taking the 2716 and 2720 CIS post-election surveys[5] as a reference (corresponding, respectively, to the 2007 Madrid regional election and the 2007 Barcelona local election)[6], an enormous number of samples from the related populations have been simulated under three different scenarios of respondent behavior during interviews and two different strategies of sampling design. Each sample has been analyzed using four alternative estimators and the accuracy of all estimates has been assessed in comparison to the actual outcomes. The results show that introducing the outcomes recorded previously in all the voting districts into the estimation process would significantly improve the accuracy of predictions.

The rest of the paper is organized as follows. The second section describes the characteristics of the target populations and the criteria followed to generate the samples. The third section describes the estimators used and the fourth analyzes and compares the forecasts. In the fifth section, the estimators are applied to the actual data collected

this perspective, if the same election could be repeated indefinitely under similar conditions (emulating in electoral terms the movie "Groundhog Day") the outcomes obtained with each new election would be different, but with some regularities that would be recognized as belonging to the same underlying population. This idea can be extended dynamically to find relationships between voting outcomes recorded at different points in time.

[5] The reason for considering post-election surveys as the basis for analysis rather than pre-election polls is due to the need to assess both the forecasting alternatives and also the sampling designs with the least possible "noise" and, above all, to be able to use an objective, external criterion of validity. The responses collected in post-election surveys tally with accomplished facts, so theoretically the estimates derived from them could be directly compared to the results actually recorded in the election. However, the answers reflected by pre-election polls are subject to statements of intent, which may vary between the date of the survey and the polling day. Consequently, some of the possible deviations observed between the values registered in the elections and predictions could be due to changes of state of opinion between the time of the survey and the polling day, rather than to technical issues. This would undoubtedly greatly hamper the assessment.

[6] Although national elections generally arouse greater interest among the public, in this study we have worked with a regional and a local poll for logistical and technical reasons. Firstly, working with a national legislative election involves dealing with more than 35,000 electoral sections (census tracts), which would have entailed enormous costs in information management. Secondly, due to the low number of representatives which are usually apportioned in most of the constituencies in a general election, national predictions are less sensitive to estimation biases and are therefore initially not as interesting in technical terms. Thirdly, from a theoretical point of view, the difficulty of producing accurate estimates grows inversely with population size and directly with the number of parties. That is, the larger the population, the less difficult (in relative terms) it is to generate precise estimates and the greater the number of candidates, the more difficult it is to get it right. The 2007 Madrid regional election was chosen as an example of a situation with a low effective number of parties, but where estimates are highly sensitive [In 2007, there were 111 representatives in the Madrid Assembly]. On the other hand, the 2007 Barcelona local elections were chosen to assess the model under theoretically more unfavorable circumstances and in a different socio-political environment. From a technical standpoint, the electorate is significantly smaller and the effective number of parties clearly higher. Politically speaking, the Catalonian electorate is more fragmented and in 2007 in Barcelona the majority party was the PSC, as opposed to Madrid where it was the PP Party.

in the 2716 and 2720 surveys. Finally, the last section summarizes the conclusions drawn and critically discusses their implications. Two appendices complete the document. The first describes how simulated vote recall has been generated, while the second provides full details of how samples and respondents' responses were simulated.

Populations, Sampling Designs and Response Simulation

According to the methodological notes accompanying the opinion polls conducted by the *Centro de Investigaciones Sociológicas* (CIS), the mechanism followed by this institution to select sample units is defined as a stratified multi-stage cluster procedure "… *con selección de las unidades primarias de muestreo (municipios) y de las unidades secundarias (secciones) de forma aleatoria proporcional, y de las unidades últimas (individuos) por rutas aleatorias y cuotas de sexo y edad*." ["…with selection of primary sampling units (municipalities) and secondary units (sections) with probability proportional to size, and of the last units (individuals) by random routes and sex and age quotas."].

In this research, the CIS sampling procedure —as described in Rodríguez Osuna (2005)—has been taken as a basis and two different strategies of sampling design have been tested. On the one hand, we have followed a similar approach to that used by the CIS (hereafter CIS; see Rodríguez Osuna, 1991 and 2005) and, on the other hand, we have simulated surveys using a significantly easier and cheaper sampling design (hereafter AL). In the case of the second strategy, a smaller number of sections have been randomly selected[7] than in the CIS samples and a relatively larger number of citizens have been surveyed in each selected section. The reason for this is to study the impact that using a theoretically cheaper sample design (due to entailing less travel expenses and pollsters) would have on the quality of the estimates.

Cluster sampling (with electoral sections as clusters) is therefore the basis of the sampling technique used in the CIS and AL surveys analyzed in this paper[8]. Having vote distributions at census tract level (i.e., the votes obtained by each party in each electoral section) is therefore essential to carry out the planned simulation. However, more data than those corresponding to the elections we intend to forecast are required in order to obtain the predictions. It is also necessary i) to have the results from the previous election in each section and ii) to associate a vote recall to each elector.[9]

Both sets of results (the votes recorded in each census tract in previous and current elections) are required to perform the analysis: past vote recall is fundamental in both the ratio estimator and post-stratification techniques currently in use and also in the methodology based on superpopulation models proposed as an alternative in this paper.

However, this requirement poses two preliminary initial problems that must be solved. On the one hand, it is necessary to establish the map of relations between the voting sections of current and previous elections and, on the other hand, it is necessary to assign a vote recall to each potential respondent. To

[7] Through random sampling without replacement, with probability of selection proportional to the size of the section and without prior stratification of the population.

[8] This comment could lead to the erroneous conclusion that the methodology proposed in this paper could not be used in telephone surveys. Nothing is further from the truth. The CIS and AL sampling designs proposed could be perfectly suited to telephone polling, although we do accept the coverage problems that telephone surveys entail. Nonetheless, coverage bias would not pose any additional concern in this framework as this could be regarded as nonresponse.

[9] In order to simplify the analysis, we decided to confine our study to the results recorded in the elections of the same type held previously in the constituency.

solve the first problem, we have employed the classical solution[10], which consists of comparing administrative codes and number of voters assigned to each section as a mechanism that establishes the chart of matches between old and new polling units by applying a set of sounded rules (Pavía-Miralles, 2005: 1117-8).[11]

Once the map of relationships among sections has been established, the marginal distributions of votes (the votes recorded in previous and current elections) are available in each section. However, we also need vote recalls (or cross distributions) for all electors to apply the proposed estimators. The vote, however, is secret, so it is not possible to know present or past voting behavior at individual level. Nevertheless, in order to assign a vote recall to each subject, one could try to exploit the aggregate information available on the voting behavior of electors to infer the electoral behavior of each type of voter in each section. This issue, known as "*the ecological inference problem*" (King, 1997), is solved in this work exploiting the valuable information on transfer voting that the 2716 and 2720 CIS electoral polls provide using, in a two-stage process, techniques of matrix balancing[12] (Pavia *et al.*, 2009). The basic idea is to estimate the voting transfer matrix among electoral options at census tract level in order to, depending on the group (electoral option) to which a subject belongs to be capable of assigning them a current vote and a distribution of vote recall conditioned on their current vote.

After applying the above process —the technical details of which are shown in an example in Appendix I— we have the information about the current and previous electoral behavior of each elector in each section. These are the populations that have been used to simulate the samples.[13]

From each of the above populations, 6,000 samples of 1,000 electors (the same sizes projected in the 2716 and 2720 surveys taken as references) were extracted. A thousand samples for each of the six scenarios obtained by combining the two proposed sampling design (CIS and AL) with the three hypotheses considered for voters' behavior when interviewed (without error, WE; with nonresponse bias, NRB; and with nonresponse bias and response error, RE; see Appendix II).

A three-step procedure was followed to simulate the samples. Census tracts were selected in the first stage. In the second sta-

[10] Currently, however, with the increasing availability of geographic information, the task of establishing correspondences could easily be automated (and improved by exploiting the spatial correlation present in election outcomes) and even extended to situations where the manual assignment is impracticable (Pavía and López-Quilez, 2012).

[11] In particular, the basic list of rules used to track and establish the relationships between census tracts of successive elections could be summarized as follows: (i) A direct match is established between polling units that have apparently not changed (under the assumption that the relatively small number of entrances and exits in their voter lists are random); (ii) When either two (or more) sections are combined to create one (or more) new section(s), the aggregate outcomes of the original sections are considered as historical data for the emerging section(s); (iii) For those new sections which stem from the division of a previously existing section, the vote proportions of the original section are assigned as their past vote proportions; and, (iv) Either neighbourhood, city or, even, constituency average vote proportions are assigned as historical data for newly (or practically newly) created polling sections, due to the fact that they are usually located in the expansion areas of the cities.

[12] In particular, the RAS method has been used. This method is a mechanical procedure, quite respectful with the initial entries of the matrix, which has been widely used in political science to infer individual behavior from aggregate values (e.g., Johnston and Pattie, 2000, Gschwend *et al.*, 2003) and that receives theoretical backing from the information theory and entropy.

[13] It should be noted that the marginal distributions of votes used in each section coincide with the real ones and although the cross-distributions are unknown, the estimates achieved should be close to reality. In any case, all the prediction strategies analyzed compete to see which generates forecasts closer to actual results over the same populations.

ge, electors were drawn[14] in each selected census tract. And, in the third stage, the current and past vote responses of the individuals selected at second stage were collected. The details about how the samples and the responses of individuals were generated are given in Appendix II.

In addition to the detailed information provided in Appendix II, we illustrate how the samples were generated using as an example the steps followed to draw electors and generate respondents' answers in an AL sample with nonresponse bias and error response of the Barcelona local election. First, we randomly selected 25 sections (with selection probability proportional to size of the section) from the 1,482 electoral sections of the Barcelona electorate to then proceed, in each selected section, as follows. An elector is chosen randomly from all the electors in the section and his/her vote is observed in order to simulate nonresponse bias by "tossing a coin" with the probability of heads and tails not necessarily being equal. Let us say that s/he votes for CiU. At this point, a coin with approximately 20% probability of *heads* (figure assigned to CiU voters)[15] is thrown. If it is heads, the voter is discarded and another elector is drawn from the citizens that have not as yet been selected. If it is tails, the voter becomes part of the sample. This process is repeated until 40 electors from the section are chosen. Then, the 40 selected voters were subjected to another "coin trial" to simulate the response error. More specifically, for each elector a new coin (with a 5% probability of heads) was tossed. If tails comes up, his/her recall and current vote is directly recorded, whereas if it is heads, his/her present and past vote are generated randomly.

Estimators

For each of the simulated samples, four alternative estimates were obtained. First, and as a reference for comparison, direct estimates (hereafter DIR) were obtained by converting the direct raw answers of respondents into percentages. This estimator is employed to ascertain the extent of the bias of each particular sample and to assess the improvement that results from incorporating vote recall into the prediction process. In addition to this simple estimator, another three estimators were used. They all seek to make more efficient use of the out-of-sample information available and use vote recall as auxiliary information to reduce (sampling and non-sampling) forecasting bias. More specifically, they are: (i) A post-stratification estimator (hereafter PS) with "individual-level correction" (Mitofsky and Murray, 2002) in the variant commonly used in Spain; (ii) A ratio-weighted estimator (hereafter RAT), with correction at constituency level[16]; and (iii) The estimator proposed as an alternative in this paper (hereafter HD) in the version presented in Pavía (2010), with corrections at section level. In all cases, estimates were constructed with the goal of predicting the percentage of valid votes that each of the main candidatures competing in the elections would achieve. The details for calculating these estimates are given below.

The direct estimator of p_j, the proportion of valid votes obtained by the *j*th party, is

[14] In all the simulated samples only resident electors have been considered as non residents clearly cannot be interviewed. Obviously, therefore, the votes of non-resident electors have not been taken into account for adjustments or comparisons either.

[15] The probability of nonresponse of each political option was determined from the information contained in the surveys that served as a reference, in this case the survey 2720. In the case of CiU voters, the probability of nonresponse was established between 15 and 25 per cent.

[16] Ratio-type estimators have a high reputation and are recommended in many books on sampling (see, e.g., Särndal *et al*., 2003). In fact, an estimator of this class was used during the 2000 US Presidential Elections to produce the exit-poll predictions (Mitofsky, 2003).

easily obtained as the ratio between the votes that party *j* receives in the poll and the number of respondents who expressed their intention of voting. That is, the direct prediction (DIR) to p_j is given by equation (1).

$$\hat{p}_{\mathrm{DIR}_j} = \frac{v_j}{v} \qquad (j = 1, 2,\ldots r), \tag{1}$$

where v_j represents the number of voters surveyed who declare they will vote for option *j*, $v(=\sum_j v_j)$ comprises the total number of individuals who have indicated they will vote in these elections and *r* is the number of options (including blank votes) concurring in the election.

In order to correct for nonresponse, coverage and measurement (response) errors, many public opinion research institutes (including CIS) typically use vote recall to weight their predictions to ensure that the sample is politically representative. Post-stratification techniques[17] figure prominently among these strategies. After grouping responses according to vote recall, the post-stratification estimator reweights each observation to guarantee that by applying the weights to the responses on past votes, previous election estimates would coincide with the actual results recorded.[18]

In particular, as the simulated samples are self-weighting, the new weights would be given by equation (2) for electors who voted in the previous election for the *j*th electoral party[19] and by equation (3) for voters who did not vote in the past call.

$$\omega_{0j} = \frac{\pi_{j,0}}{\frac{v_{j,0}}{n}}, \qquad (j = 1, 2,\ldots, r) \tag{2}$$

$$\omega_{0,r+1} = \frac{1 - \sum_{h=1}^{r} \pi_{h,0}}{\frac{n - v_0}{n}}, \tag{3}$$

where $v_{j,0}$ represents the number of voters who said they voted for option *j* in the previous election, v_0 $(=\sum_j v_{j,0})$ denotes the total number of electors in the survey who voted in the preceding election, $\varpi_{j,0}$ represents the proportion of votes (over census) recorded for option *j* in the previous election and *n* is the effective sample size.

The weights obtained in (2) and (3) are used as elevation factors to obtain new estimates. Each individual is weighted according to his/her vote recall. More specifically, if $v_{i,j}$ is the number of electors in the poll who choose option *j* in the current election (where $v_j = \sum_i v_{i,j}$) having declared to have chosen option *i* (i = 1.2, ..., r +1) in the previous election, the reweighted number of voters for option *j* in the current election is obtained through equation (4) and, hence, through equation (5), the PS predictions for the proportions of valid vote of each party[20].

$$\hat{v}_j = \sum_{h=1}^{r+1} \omega_{0i}\, v_{i,j} \qquad (j = 1, 2,\ldots, r) \tag{4}$$

$$\hat{p}_{\mathrm{PS}j} = \frac{\hat{v}_j}{\sum_{h=1}^{r} \hat{v}_h}, \qquad (j = 1, 2,\ldots, r) \tag{5}$$

As an alternative to PS forecasts (which are acquired taking into account the indivi-

[17] In the presence of nonresponse, however, according to Kalton and Kasprzyk (1986), the commonly called post-stratification estimator should be denoted "*the population weighting adjustment estimator*".

[18] This technique, therefore, could also be seen as a particular case of a calibration procedure, where past outcomes are used as an auxiliary (calibration) variable (e.g., Särndal, 2007).

[19] To make the notation less dense, it is assumed that the same number of parties competed in both elections.

[20] The same predictions (although with different estimation error) could also have been reached by adopting a superpopulation scheme in which the individual probabilities of change among political options were modeled exclusively depending on the previous vote (Aybar, 1998).

dual responses of all subjects participating in the survey grouped into strata), estimates could also be obtained using a correction at constituency level in the spirit of the ratio predictors (see, e.g., Särndal et al., 2003: 180) with the initial forecasts being adjusted using vote recall as a covariate. In this case predictions are achieved using two separate ratio estimates derived from the survey data and after applying a final correction to remove the inconsistency that the well-known lack of unbiasedness of the ratio estimator causes in the forecasts.

In particular, if $p_{j,0}$ denotes the proportion of valid votes attained by party *j* in the previous election and $\hat{p}_{\mathrm{DIR},j,0}$ represents the poll direct estimate for the proportion of votes gained by party *j* in the previous election (calculated as the ratio between those respondents who declared in the survey they had voted for party *j* in the past call and the total respondents who said they voted in previous elections), the RAT predictions are obtained by applying equations (6) and (7) recursively. First, through equation (6), an initial prediction for the proportion of vote for party *j*, $\tilde{p}_j$, is obtained using the classical ratio estimator. Then, in a second stage, we obtain the final RAT estimates by applying expression (7), which re-weights all the individual estimates given by the $\tilde{p}_j$, estimates to sum to one. This second stage overcomes the main drawback of the classical ratio estimator, which almost certainly yields a set of predictions the sum of which is not unitary.[21]

$$\tilde{p}_j = \hat{p}_{\mathrm{DIR},j} \frac{p_{j,0}}{\hat{p}_{\mathrm{DIR},j,0}}, \qquad (j = 1, 2,\ldots, r) \quad (6)$$

$$\hat{p}_{\mathrm{RAT},j} = \frac{\tilde{p}_j}{\sum_{h=1}^{r} \tilde{p}_h}, \qquad (j = 1, 2,\ldots, r) \quad (7)$$

Finally, as an alternative to RAT and PS estimates, the HD[22] superpopulation estimator, which was the real motivation for this research, was proposed. In fact, the inspiration of this research was i) to study the capacity of the HD estimator to correct nonresponse bias and ii) to compare its performance against both PS and RAT estimators in order to assess whether the theoretically more intensive use that it makes of available information is worthwhile. As well as employing the individual responses of all respondents to produce forecasts, the HD estimator also considers in which census tract each respondent is enrolled in and, moreover, exploits the historical data of all the electoral sections through a superpopulation model, rather than just their aggregation, as is the case with PS and RAT estimators.

This study uses the version of the HD estimator proposed by Pavía (2010), who follows a multistage procedure with different forecasting strategies for large and small parties[23]. First, in each of the sections sampled, initial predictions corrected by vote recall are obtained for the proportion of votes that would reach each of the major parties. Second, using the congruence that electoral results from consecutive elections display in small area bases, the proportions of votes in both non-sampled and sampled sections are predicted by regressing the estimates obtained in the sampled sections in step one on the outcomes recorded in these same sections in the previous elections (Pavía-Miralles, 2005). In the third place, all the section forecasts obtained in step two are added to generate an estimate at constituency level for large parties [PP and PSOE were considered

[21] Following Särndal *et al.* (2003: 180), if recall vote estimates are much skewed, the bias can even be awfully significant.

[22] The abbreviation HD (coined in Pavía *et al.*, 2008) denotes *historical data*. The estimator is named after the intensive use of historical data for the small areas it is based on.

[23] Pavia (2010: 73) justifies taking a different approach for large and small parties on the basis of the law of large numbers.

as the major political parties in the case of the Madrid regional elections and PSC, CiU and PP in the Barcelona local elections]. Finally, HD predictions for smaller parties are attained by combining the forecasts obtained for the major parties with the direct estimates, $\hat{p}_{\mathrm{DIR},j}$, of small parties.

In particular, adapted to the current situation, the *modus operandi* of the process proposed in Pavía (2010) would operate as follows:

i) From the survey data, initial estimates for the proportions of current and past votes are obtained in each electoral sampled section s (=1,2,…,N_s) and for each large party j (=1.2,..,G).

$$\tilde{p}_{js} = \frac{v_{js}}{v_s}, \quad \tilde{p}_{js,0} = \frac{v_{js,0}}{v_{s,0}} \qquad (8)$$

$$(s = 1, 2,\ldots,N_s \text{ y } j = 1, 2,\ldots, G),$$

where v_{js} ($v_{js,0}$) denotes the number of respondents in the section s who declare a vote for party j in the current (previous) election, $v_s = \sum_{j=1}^{r} v_{js}$ ($v_{s,0}$) represents the total respondents in the section who declare a vote in the current (previous) election and N_s is the number of sections sampled.

ii) Using actual values recorded in the previous election in the sampled sections (and assuming that entrances and exits in section voter lists are random), vote recall adjusted estimates, $\ddot{p}_{js}$, are obtained for the proportions of votes that each party j would attain in each section s by:[24]

$$\ddot{p}_{js} = \tilde{p}_{js} + (p_{js0} - \tilde{p}_{js0}),$$

$$(s = 1, 2,\ldots, N_s \text{ y } j = 1, 2,\ldots, G), \qquad (9)$$

where $p_{js,0}$ denotes the proportion of votes recorded in section s by party j in previous elections.

iii) There is assumed to be a linear relationship between the proportion of actual (p_{js}) and previous votes ($p_{js,0}$) for each party at section level, with (for simplicity) zero mean normal disturbances with constant correlation between parties and conditional independence between sections; i.e.:

$$p_{js} = \alpha_j + \beta_j p_{js,0} + \varepsilon_{js},$$

$$(s = 1, 2,\ldots, \text{N y } j = 1, 2,\ldots, r), \qquad 10)$$

ε_{js} being 0-mean normal disturbances verifying $E(\varepsilon_{js},\varepsilon_{j^*s^*}) = \delta_{ss^*}\sigma_{jj^*}$ (where δ is Kronecher's delta function), α_j and β_j unknown parameters and N the number of sections in the constituency.

iv) Using the large party predictions obtained at section level in ii) and defining $\ddot{p}_{G+1,s} = 1 - \sum_{j=1}^{G} \ddot{p}_{js}$ as the estimated proportion of votes gained for the remaining options (OT) in section s, the parameters α_j and β_j (for j = 1,2,…,G+1) of (10) are estimated via the iterative algorithm proposed in Pavía-Miralles (2005, 1121)[25] and conditioned on these parameter estimates, predictions for the proportion of votes gained by the major parties in each section are reached through:

$$\hat{p}_{js} = \hat{\alpha}_j + \hat{\beta}_j\, p_{js,0},$$

$$(s = 1, 2,\ldots, \text{N y } j = 1, 2,\ldots, \text{G}), \qquad (11)$$

v) Once estimates for the major parties are available in all census tracts, the section forecasts are added up to reach constituency HD estimates, using equation (12) for large

24 Correcting nonresponse bias at section level makes the process more flexible, due to as Pavía (2010: 70) points out, "*[i]t allows for a different bias mechanism for each polling place and for the magnitude and even the direction of the bias to vary across locations.*"

25 In order to simplify the estimation process, it is assumed that errors in measuring the dependent variable are absorbed by the error term (see, Greene 2003: 84). Pavía-Miralles and Larraz-Iribas (2008) offer an alternative algorithm to deal with this issue when the measurement errors in the dependent variable are considered explicitly.

parties and equation (13) for small candidatures:

$$\hat{p}_{\mathrm{HD},j} = \sum_{s=1}^{N} \omega_s \, \hat{p}_{js}, \quad (j = 1, 2,\ldots, \mathrm{G}), \quad (12)$$

$$\hat{p}_{\mathrm{HD},j} = \frac{\hat{p}_{\mathrm{DIR},j}\left(1 - \sum_{j=1}^{\mathrm{G}} \hat{p}_{\mathrm{HD},j}\right)}{\sum_{j=1+1}^{r} \hat{p}_{\mathrm{DIR},j}},$$

$$(j = \mathrm{G} + 1,\ldots, r) \quad (13)$$

where ω_s is the weight of the sth section in the constituency, defined by the product of the participation rate recorded in the section in the previous elections and the number of electors in the section in the current elections:

$$\omega_s = \frac{e_s t_{s,0}}{\sum_{h=1}^{N} e_h t_{h,0}}, \qquad (s = 1, 2,\ldots, \mathrm{N}) \quad (14)$$

e_s being the number of voters in the sth section and $t_{s,0}$ the participation rate recorded in the sth section in the previous elections.

Analysis of Simulations

According to statements made in the previous sections, using Monte Carlo techniques[26], a large number of samples have been simulated under different scenarios in order to assess, as neutrally as possible, the capacity displayed by each one of the proposed estimators to correct the incidence of nonresponse bias in forecasting. One thousand samples were generated in each scenario and four estimates of each sample were obtained. This implies 4,000 estimates for each political party and scenario, or equivalently, 24,000 sets of estimates for each election. Hence, given the vast number of available estimates, it is necessary to use a statistical summary of the results to draw general conclusions. In order to do so, we have pursued a multilateral approach considering both the accuracy of the estimates for each party and the degree of overall fit between estimates and actual values. Thus, in addition to computing the usual summary statistics (mean, median, standard deviation, first and third quartiles, minimum, maximum, coefficients of variation, skewness and kurtosis) to assess and compare isolated party prediction and error estimate distributions, we have also evaluated the overall level of adjustment shown between sets of estimates and actual outcomes. The full array of error measures listed in Table I, which also details the mathematical expressions of the adjustment statistics used, were used for the foregoing adjustment. Comparisons of the distributions of these statistics in each scenario, calculated over all the samples, have been employed as a basis for assessing the estimators analyzed.

In order to lighten the notation, a coding system based on the acronyms introduced in the preceding sections —DIR, PS, RAT and HD to mark estimators, CIS and AL to indicate sample design strategy and WR, NRB and RE to refer to assumptions about voter behavior when interviewed— have been used in the results shown in the tables and figures that follow.

Madrid Regional Elections

This subsection analyzes and presents the main results of the simulations obtained for the 2007 Madrid Assembly regional elections. Of all the results, those corresponding to the samples generated in scenarios with nonresponse bias and response error should be more carefully observed, due to being tho-

[26] Monte Carlo methods include a wide range of techniques that, using repeated random experiments, seek to find answers to problems that cannot normally be overcome analytically.

TABLE I. *Error measures used to assess goodness-of-fit of forecasts*

Description	Acronyms	Equations(1)
Mean Square Error	MSE	$\frac{1}{r}\sum_{k}(p_k - \hat{p}_{x,k})^2$
Root Mean Square Error	RMSE	$\sqrt{\frac{1}{r}\sum_{k}(p_k - \hat{p}_{x,k})^2}$
Absolute Mean Error	AME	$\frac{1}{r}\sum_{k}\lvert p_k - \hat{p}_{x,k}\rvert$
Relative Mean Error	RME	$\frac{100}{r}\sum_{k}\frac{\lvert p_k - \hat{p}_{x,k}\rvert}{p_k}$
Entropy	ENT	$-\sum_{k} p_k \log\left(1 - \left\lvert\frac{p_k - \hat{p}_{x,k}}{100}\right\rvert\right)$

Source: Own elaboration.

(1) r denotes the number of policy options for which the joint adjustment is calculated, p_k the actual percentage of votes recorded for the *k*th option, and $\hat{p}_{x,k}$ the estimate of the percentage of votes gained by party *k* after applying the corresponding estimator, with X = DIR RAT, PS and HD.

se closest to real conditions. A broad summary of several relevant aspects that are important to analyze the simulations are provided in Tables II and III and in Figure 1. Table II, classified by scenario and estimator, includes forecast means and associated biases. The average values of the goodness-of-fit measures and a comparison of the distributions of the degree of adjustment of all the estimates, measured by the percentage of times that each estimator generates the best solution in terms of entropy, are displayed in Table III. Finally, the box and whisker plots in Figure 1 show the distributions of the predictions obtained following the CIS sampling design.

Of the two elements that define each scenario, the willingness of voters to participate in the survey is the question that has the greatest impact on estimate accuracy. In fact, the impact of sampling design can even be classified as minor. This is not innocuous, however. On average, the estimates achieved using the CIS design are slightly closer to actual values than those obtained with the AL strategy, although both sets of predictions show fairly comparable fit levels. The reduction in the number of sampling points (and in their spatial distribution) that the AL design involves (compared to CIS plan) leads, on the one hand, to a petty increase in the bias estimation of the percentages for major parties (see Table II) and, on the other hand (as expected), a slight increase in the variability of predictions. Nonetheless, these changes do not apply equally to all estimators. The HD estimator suffers the least, in terms of sample unbiasedness and variability[27], a change in the sampling plan, and improves its relative position in regard to its competitors —PS and RAT estimators, see Table III.

The impact on forecasts of the assumptions about the behavior of voters when polled, however, is more evident. Under ideal conditions (i.e., when samples are generated without error) all the estimators produce, as

27 In fact, when nonresponse bias is present, the HD estimator displays even lower levels of variability.

TABLE II. *Forecast and estimation error averages(1) of the percentages of votes for the main parties contesting the 2007 Madrid Assembly regional election*

Escenario	Estimador	Percentages				Errors(3)			
		PP	PSOE	IU	OT	PP	PSOE	IU	OT
CIS_WE	DIR	53.43	33.13	8.80	4.65	0.29	–0.21	–0.07	–0.02
	PS	53.12	33.32	8.95	4.61	–0.02	–0.01	0.09	–0.06
	RAT	52.78	33.35	9.03	4.84	–0.36	0.01	0.17	0.17
	HD	53.16	33.15	8.94	4.75	0.02	–0.18	0.08	0.08
AL_WE	DIR	52.86	33.67	8.84	4.62	–0.28	0.34	–0.02	–0.04
	PS	52.89	33.65	8.85	4.61	–0.25	0.32	–0.01	–0.06
	RAT	52.71	33.59	8.88	4.82	–0.43	0.26	0.01	0.16
	HD	53.03	33.46	8.85	4.66	–0.11	0.13	–0.01	–0.01
CIS_NRB	DIR	51.13	34.72	9.30	4.84	–2.01	1.39	0.44	0.18
	PS	51.56	34.51	8.95	4.97	–1.58	1.18	0.09	0.31
	RAT	51.65	34.33	8.81	5.21	–1.49	0.99	–0.05	0.54
	HD	52.02	34.09	9.12	4.77	–1.12	0.76	0.26	0.10
AL_NRB	DIR	50.95	34.86	9.36	4.83	–2.19	1.52	0.49	0.17
	PS	51.54	34.61	8.88	4.97	–1.60	1.28	0.02	0.31
	RAT	51.75	34.43	8.67	5.16	–1.39	1.09	–0.20	0.49
	HD	52.11	34.22	8.99	4.68	–1.02	0.88	0.13	0.01
CIS_RE	DIR	49.99	34.08	10.11	5.83	–3.15	0.74	1.24	1.16
	PS	50.78	34.18	9.33	5.70	–2.35	0.84	0.47	1.04
	RAT	51.99	34.54	8.73	4.74	–1.15	1.20	–0.13	0.08
	HD	52.12	34.34	8.58	4.97	–1.02	1.00	–0.28	0.30
AL_RE	DIR	49.70	34.42	10.13	5.76	–3.44	1.08	1.27	1.09
	PS	50.69	34.41	9.27	5.64	–2.45	1.07	0.41	0.97
	RAT	51.98	34.70	8.61	4.71	–1.16	1.37	–0.25	0.04
	HD	52.07	34.54	8.52	4.87	–1.06	1.21	–0.35	0.20
Eleccion	Outcomes(2)	**53.14**	**33.33**	**8.86**	**4.67**	—	—	—	—

Source: Own elaboration.

(1) Mean values from 1,000 simulated samples.

(2) Percentage of valid votes recorded in the resident population.

(3) Computed as the average percentage difference between estimated and real values.

expected, highly accurate predictions, with a negligible average bias. However, generally speaking we can say that the PS estimator registers the best fit supported largely by the fact it produces the best predictions for small candidatures followed by the HD and RAT predictors (with similar figures) and the DIR estimator, which clearly shows superior levels of error. With CIS design, nevertheless, the DIR estimator generates the best solution slightly more times than HD and RAT estimators (see Table III), which combined with the above statement suggests it is less robust (see Figure 1). That is, when it errs, it errs by more. In short, we can say that even in circumstances where direct estimation has no theoretical disadvantage, vote recall correction helps to improve estimate accuracy.

Voters' behavior simulated in the ideal scenarios, however, is far from realistic. In the presence of asymmetric nonresponse, all four estimators are hit hard. DIR estimates show a significant bias and although the use of the auxiliary information provided by vote recall substantially reduces the magnitude of the bias, it remains appreciable. Notwithstanding, the HD strategy does reduce bias the

TABLE III. *Summary of goodness-of-fit measures between actual and estimated joint vote distributions. 2007 Madrid Assembly regional elections*

Scenario	Estimador	% success over 1000 simulations(1)				Goodness-of-fit-measures(2)(3)				
		% success	PS *v* HD	PS *v* RAT	RAT *v* HD	ENT	MSE	RMSE	AME	RME
CIS_WE	DIR	24.0	—	—	—	1.62	2.70	1.47	1.24	8.12
	PS	33.0	65.3	64.7	—	1.07	1.35	1.05	0.90	6.81
	RAT	22.0	—	35.3	53.7	1.30	2.08	1.31	1.12	9.20
	HD	21.0	34.7	—	46.3	1.36	2.13	1.32	1.12	8.33
AL_WE	DIR	18.4	—	—	—	2.38	5.62	2.03	1.68	9.31
	PS	35.0	60.7	58.7	—	1.23	1.69	1.16	0.99	6.99
	RAT	23.0	—	41.3	49.2	1.41	2.29	1.37	1.17	9.08
	HD	23.6	39.3	—	50.8	1.41	2.22	1.35	1.15	8.20
CIS_NRB	DIR	25.1	—	—	—	2.05	4.01	1.76	1.46	8.37
	PS	18.8	43.8	51.7	—	1.50	2.20	1.33	1.12	7.20
	RAT	21.2	—	48.3	39.2	1.50	2.48	1.41	1.20	9.01
	HD	34.9	56.2	—	60.8	1.38	2.06	1.28	1.08	7.41
AL_NRB	DIR	22.5	—	—	—	2.68	6.73	2.22	1.83	9.72
	PS	22.1	42.6	50.4	—	1.62	2.50	1.42	1.20	7.56
	RAT	24.8	—	49.6	39.3	1.59	2.74	1.49	1.27	9.35
	HD	30.6	57.4	—	60.7	1.44	2.20	1.33	1.13	7.79
CIS_RE	DIR	16.4	—	—	—	2.51	5.74	2.21	1.87	13.12
	PS	12.1	30.2	29.5	—	1.87	3.19	1.66	1.40	10.25
	RAT	34.0	—	70.5	49.1	1.48	2.35	1.38	1.17	8.16
	HD	37.5	69.8	—	50.9	1.46	2.31	1.36	1.15	7.91
AL_RE	DIR	15.1	—	—	—	3.08	8.68	2.62	2.20	13.93
	PS	13.0	27.4	31.3	—	2.01	3.66	1.75	1.49	10.32
	RAT	35.4	—	68.7	45.0	1.56	2.62	1.45	1.23	8.36
	HD	36.5	72.6	—	55.0	1.51	2.47	1.41	1.20	8.11

Source: Own elaboration.

(1) Percentage of samples for which the corresponding estimator achieves a better fit in terms of entropy.

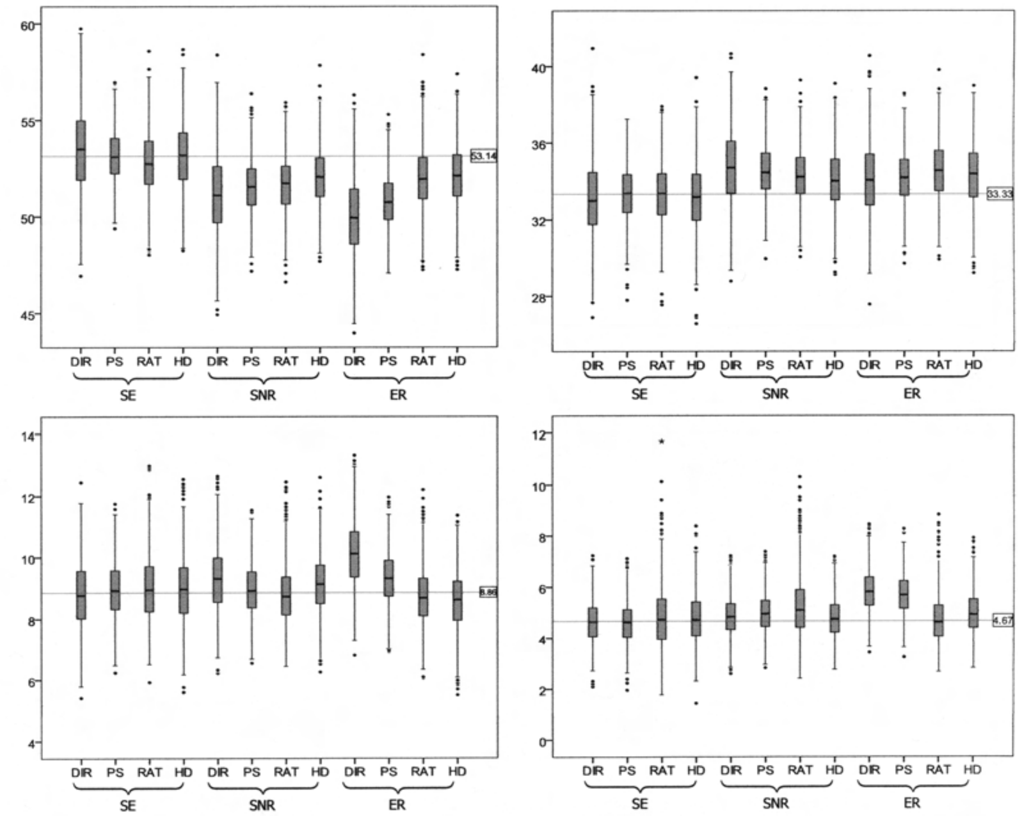
(2) Mean values from 1,000 simulated samples.

(3) ENT: Entropy; MSE: Mean square error; RMSE: Root of MSE; AME: Absolute mean error; RME: Relative mean error.

most and records the best scores in terms of forecasting accuracy. Indeed, it could be argued that the larger the bias of the sample, the better the HD estimator works (in relative terms). In fact, while the HD predictor is as a rule the best (slightly more than) a third of the times (see Table III), we found that this figure rises to almost 60% when considering exclusively the hundred samples with the most bias. The HD estimator also triumphs in pairwise comparisons reversing the trend observed with ideal samples after improving its figures against the RAT predictor and especially by advancing in comparison to the PS estimator. In fact, the relatively low variability showed by PS estimates together with their persistent bias, indicates that the PS estimator has the lowest success rates in these scenarios.

The picture is quite similar in RE scenarios, in spite of in comparison to NRB settings RE samples see their average levels of unbiasedness decrease (albeit minimally) for all estimators. Although the convergence that occurs between HD and RAT estimators stands out, the order of preference (HD,

FIGURE 1. *Box and Whisker plots of all the estimates obtained using CIS sampling design for PP (top left), PSOE (top right), IU (bottom left) and others (bottom right).*



RAT, PS and DIR) obtained in NRB scenarios remains valid. In terms of overall adjustment, the PS predictor once again suffers the most (see Table III). In fact, in the race of the four estimators, it only works better in just over 10% of cases and drops back to odds of 7 to 3 in the pairwise comparisons against HD and RAT estimators. The post-stratification strategy encountered the most serious problems when predicting the two major parties, with the largest errors concentrated in PP estimates. In light of the results, it can be argued that the PS estimator corrects raw responses the least out of the three estimators based on vote recall and also yields the predictions that are closest to DIR estimates.

Nevertheless, despite the HD estimator being the best in both RE and NRW scenarios and the fact that the confidence intervals of HD predictions[28] had included the true value in a large percentage of cases and that consequently there would have not have been errors but uncertainty, from a statistical standpoint, average HD results are still a significant distance from true values. Consequently,

[28] Obtained after taking the variance of the 1,000 simulations as an estimate of the sampling variance.

there is still room to search for more accurate estimators.

Barcelona Local Elections

The simulation outcomes obtained for the 2007 Barcelona local elections are discussed in this subsection. Compared to the Madrid Assembly elections, the electoral scene in Barcelona has an evident multiparty structure and a considerably smaller electorate. Both factors make it more difficult to obtain accurate estimates. Tables IV and V are the namesakes of Tables II and III in the previous subsection, while Figure 2 in this case provides the estimate distributions obtained after applying the AL sampling design.

Barcelona election simulation results reinforce the findings reached using data from the Madrid elections. Despite the situation being more complex, the electorate being smaller in size and the political landscape more fragmented, the general trends and overall results obtained previously are confir-

TABLE IV. *Forecast and estimation error averages[1] of the percentages of votes for the main parties contesting the 2007 Barcelona local elections*

Scenario	Estimator	Percentages						Errors[3]					
		PSC	CiU	PP	ICV	ERC	OT	PSC	CiU	PP	ICV	ERC	OT
CIS_WE	DIR	30.10	25.33	15.58	9.36	8.84	10.79	0.14	–0.13	–0.03	0.01	–0.01	0.02
	PS	30.19	25.40	15.78	9.17	8.87	10.59	0.23	–0.05	0.17	–0.17	0.02	–0.20
	RAT	30.05	25.29	15.68	9.34	8.74	10.90	0.09	–0.16	0.07	–0.01	–0.11	0.12
	HD	30.18	25.04	15.64	9.41	8.88	10.85	0.22	–0.42	0.03	0.06	0.03	0.08
AL_WE	DIR	29.97	25.30	15.64	9.49	8.81	10.79	0.01	–0.16	0.04	0.15	–0.04	0.00
	PS	30.06	25.47	15.82	9.24	8.85	10.56	0.10	0.01	0.22	–0.11	–0.01	–0.21
	RAT	29.93	25.44	15.72	9.36	8.74	10.81	–0.03	–0.02	0.12	0.01	–0.11	0.03
	HD	30.17	25.19	15.68	9.45	8.76	10.75	0.21	–0.26	0.08	0.10	–0.10	–0.03
CIS_NRB	DIR	33.99	26.46	9.20	11.46	9.11	9.78	4.03	1.01	–6.41	2.12	0.26	–1.01
	PS	31.27	25.11	13.01	11.87	8.63	10.11	1.31	–0.35	–2.59	2.52	–0.22	–0.67
	RAT	30.07	24.62	14.14	11.31	8.73	11.13	0.11	–0.83	–1.47	1.96	–0.13	0.36
	HD	30.02	24.84	14.85	11.44	9.08	9.77	0.06	–0.61	–0.75	2.09	0.23	–1.02
AL_NRB	DIR	34.14	26.36	9.16	11.56	9.00	9.78	4.19	0.91	–6.45	2.22	0.15	–1.02
	PS	31.38	25.10	12.93	11.95	8.55	10.09	1.42	–0.36	–2.68	2.61	–0.30	–0.69
	RAT	30.17	24.66	14.05	11.38	8.66	11.08	0.21	–0.79	–1.55	2.03	–0.19	0.29
	HD	30.11	24.83	14.76	11.55	8.98	9.77	0.15	–0.63	–0.85	2.20	0.13	–1.00
CIS_RE	DIR	32.80	25.49	9.45	11.72	9.45	11.09	2.84	0.04	–6.15	2.37	0.60	0.30
	PS	30.99	24.69	12.71	12.00	8.92	10.69	1.03	–0.77	–2.90	2.66	0.07	–0.09
	RAT	30.80	24.99	14.58	11.68	9.19	8.76	0.84	–0.47	–1.02	2.33	0.34	–2.02
	HD	30.18	24.60	14.92	11.01	8.87	10.42	0.22	–0.86	–0.69	1.66	0.02	–0.35
AL_RE	DIR	32.96	25.55	9.44	11.62	9.43	11.00	3.00	0.10	–6.17	2.27	0.58	0.22
	PS	31.05	24.75	12.67	11.96	8.92	10.65	1.09	–0.71	–2.93	2.61	0.07	–0.13
	RAT	30.78	25.03	14.56	11.66	9.21	8.76	0.82	–0.43	–1.04	2.31	0.36	–2.02
	HD	30.22	24.58	14.91	10.98	8.90	10.41	0.26	–0.87	–0.70	1.63	0.05	–0.37
Eleccion	Outcomes[b]	**29.96**	**25.46**	**15.61**	**9.35**	**8.85**	**10.77**	—	—	—	—	—	—

Source: Own elaboration.

(1) Mean values from 1,000 simulated samples.

(2) Percentage of valid votes recorded in the resident population.

(3) Computed as the average percentage difference between estimated and real values.

TABLE V. *Summary of goodness-of-fit measures between actual and estimated joint vote distributions. 2007 Barcelona local elections*

Scenario	Estimator	% success over 1.000 simulations(1)				Goodness-of-fit measures(2)(3)				
		% success	PS *v* HD	PS *v* RAT	RAT *v* HD	ENT	MSE	RMSE	AME	RME
CIS_WE	DIR	20.3	—	—	—	1.53	3.07	1.65	1.35	9.14
	PS	35.3	72.0	66.0	—	1.10	1.69	1.23	1.02	7.19
	RAT	25.0	—	34.0	58.0	1.26	2.39	1.45	1.19	8.56
	HD	19.4	28.0	—	42.0	1.34	2.43	1.47	1.22	8.49
AL_WE	DIR	16.8	—	—	—	1.80	4.14	1.89	1.55	10.20
	PS	35.8	63.9	63.8	—	1.15	1.81	1.27	1.06	7.51
	RAT	24.0	—	36.2	48.1	1.30	2.51	1.49	1.23	8.78
	HD	23.4	36.1	—	51.9	1.29	2.29	1.43	1.19	8.43
CIS_NRB	DIR	0.2	—	—	—	3.18	13.03	3.56	2.77	17.86
	PS	17.4	28.0	28.0	—	1.56	3.82	1.90	1.57	11.73
	RAT	38.1	—	72.0	46.0	1.36	3.15	1.68	1.39	10.56
	HD	44.3	72.0	—	54.0	1.30	2.69	1.57	1.31	10.04
AL_NRB	DIR	0.3	—	—	—	3.43	14.73	3.76	2.95	18.81
	PS	15.2	22.9	27.9	—	1.69	4.29	2.01	1.68	12.37
	RAT	33.8	—	72.1	39.0	1.46	3.49	1.77	1.48	11.13
	HD	50.7	77.1	—	61.0	1.34	2.91	1.63	1.35	10.42
CIS_RE	DIR	1.4	—	—	—	2.71	10.91	3.25	2.50	16.93
	PS	10.9	19.0	33.0	—	1.62	4.23	2.00	1.64	12.18
	RAT	24.2	—	67.0	30.0	1.47	3.70	1.85	1.54	12.07
	HD	63.5	81.0	—	70.0	1.25	2.31	1.45	1.21	8.91
AL_RE	DIR	0.8	—	—	—	3.00	12.29	3.43	2.67	17.54
	PS	13.6	18.3	35.0	—	1.67	4.39	2.03	1.67	12.32
	RAT	19.9	—	65.0	24.2	1.51	3.80	1.87	1.57	12.25
	HD	65.7	81.7	—	75.8	1.25	2.31	1.44	1.20	8.86

Source: Own elaboration.

(1) Percentage of samples for which the corresponding estimator achieves a better fit in terms of entropy.

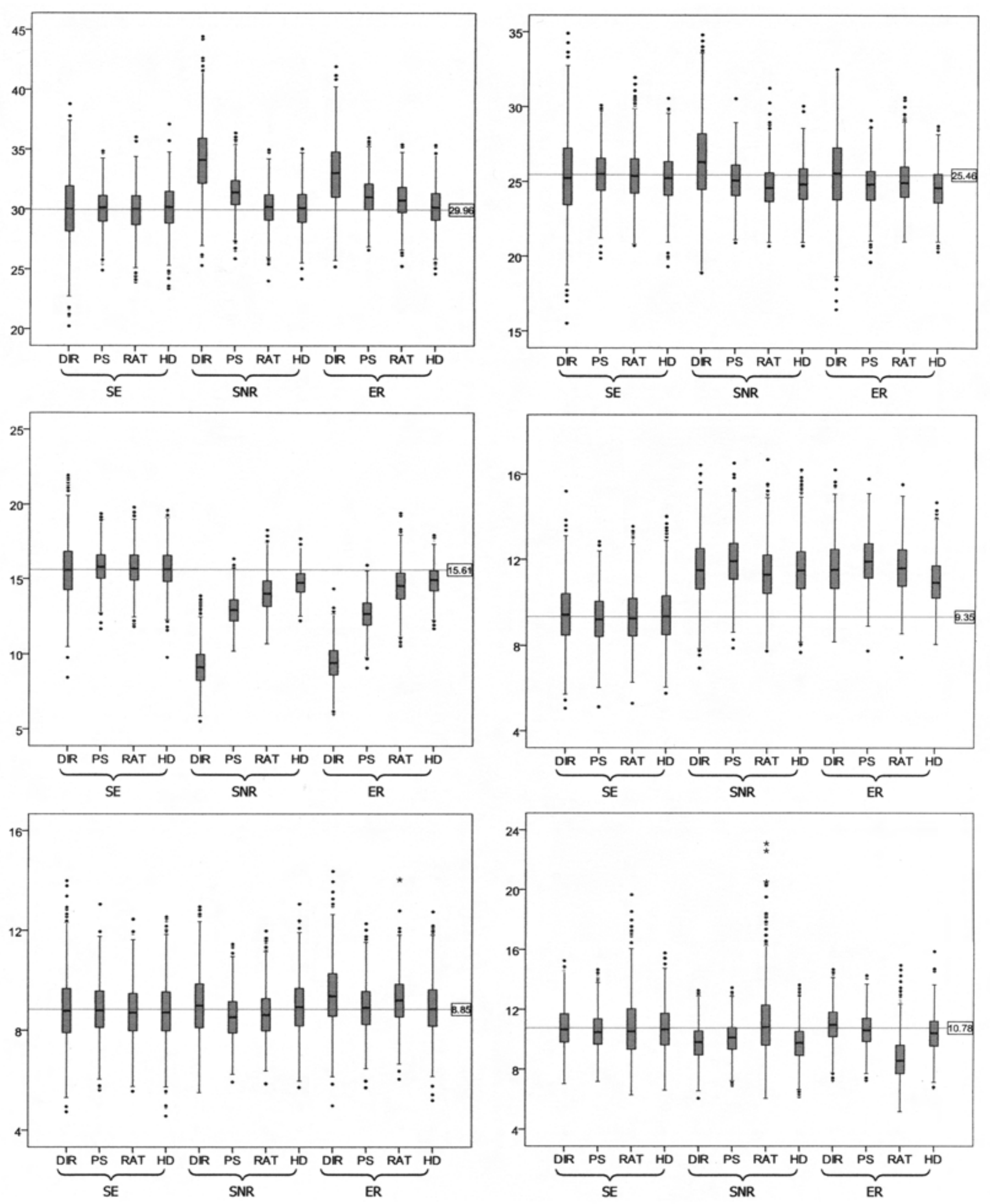
(2) Mean values from 1,000 simulated samples.

(3) ENT: Entropy; MSE: Mean square error; RMSE: Root of MSE; AME: Absolute mean error; RME: Relative mean error.

med. As regards comparisons between CIS and AL sampling, similar conclusions to those attained with the Madrid simulations are reached again. In general, selecting a smaller number of sampling points (combined with interviewing a larger number of electors at each sampled point) hardly affects the accuracy of the estimates, with the HD estimator suffering the least (if at all) the sampling strategy shift. Therefore, it could stated that financial criteria (and issues related to the estimation accuracy of other variables also surveyed) should be weighed up in order to decide, by way of a cost-benefit analysis, whether it is better to follow a CIS sampling plan or a simple sampling plan such as the AL design analyzed in this study, especially if the estimator adopted is the HD.

Likewise, focusing our attention now on the impact that voters' behavior has on predictions, we found that the results are again in line with those achieved in the previous subsection. The PS predictor once again generates the best forecasts under ideal conditions. This, however, does not disable its competitors, due, as in the case of Madrid, to

FIGURE 2. *Box and Whisker plots of all the estimates obtained using AL sampling design for PSC (top left), CiU (top right), PP (middle left), ICV (middle right), ERC (bottom left) and others (bottom right)*



the predictions obtained using this predictor also being highly accurate (see Tables IV and V)[29]. On the other hand, the HD estimator emerges again as the best when the samples have been simulated with nonresponse bias and response error. In this case, moreover, its comparative advantage against the RAT estimator increases substantially, while the superiority over the PS estimator is already huge. Most likely, the growth in terms of the relative advantage experienced by the HD estimator is due to the fact that, in this case, we must tackle significantly more biased samples, an issue that, as was discussed previously, seems to favor HD estimates. Indeed, as can be observed in Table IV, this seems to be the cause —in NRW and RE scenarios, DIR predictions display huge differences in bias, even exceeding six points on average in the case of the PP.

As regards the parties themselves, it is worth highlighting the forecasts obtained for ICV and CiU, on the negative side, and the estimates achieved for PSC and PP on the positive side. ICV accumulates the highest prediction errors. None of the estimators used seem to have been able to significantly reduce the bias present in the raw data for this party (see Figure 2). The case of CiU is somewhat different. CiU tends to display positive bias in samples simulated with nonresponse bias. Although this appears to be detected by all the strategies, the fact is that all of them over-correct it, resulting on average in significantly negatively biased predictions. In both cases, in order to gain insight into the causes of the situations described above, it would be interesting to study the spatial distribution of their support and to analyze whether there were any specific anomalies in voting transfer among parties between the 2003 and 2007 elections. At the other end of the scale it is worth highlighting the predictions achieved for PSC and PP. In this case, despite having raw data spectacularly biased, all estimators are observed to have made important corrections, which in the case of the HD predictor led to fairly accurate forecasts.

Finally, it should be noted that despite the superiority of the HD estimates being obvious in the most realistic circumstances and the fact that their average forecasts are also now relatively closer to actual outcomes, the observations made in the previous paragraph confirm that there is still room for improvement and researchers should search for more accurate strategies. In any case, it seems evident that in the presence of asymmetric nonresponse, the HD strategy clearly outperforms the RAT and PS strategies, which are the ones currently in use in the electoral polling industry.

Estimating with real data: 2716 and 2720 CIS polls

The conclusions of the simulation exercise presented in the previous section show that, in the presence of nonresponse bias, the HD estimator generates the best estimates for vote distribution more frequently. This section provides the predictions that would be obtained if the four estimators discussed in this research were applied to the actual data collected in the surveys used as a reference in this study.

So far we have considered only one type of nonresponse: Total. In real surveys, however, is very common to observe also another type of nonresponse: Partial. That is, to find individuals that have only answered to some of the posed questions. In these situations, analysts must decide whether to: i) exclude the individuals for whom there is no response in all relevant variables from the analysis or; ii) adopt a theoretically more efficient approach and use all the available sample information to predict (impute) the missing values of the relevant variables.

[29] In fact, RAT estimates work even better on this occasion than PS estimates in terms of bias.

In our case, the relevant responses to forecast are the answers to the questions on actual voting and vote recall. The 2716 and 2720 CIS surveys, which were intended to have a sample size 1,000, effectively had 969 and 974 observations, respectively. However, only 807 and 728 people surveyed gave respectively a valid response to the two items required to apply vote-recall based strategies. A few more respondents provided feedback at least about their actual vote: 881 and 826, respectively. In light of these data, we have considered three different scenarios in which to use the data collected in the surveys to yield predictions. In a first stage, only the raw data corresponding to individuals who report their current and previous vote are used. This approach however is not theoretically efficient as it makes no use of a lot of information still available in the survey. Thus, we have considered two additional scenarios in which the relevant responses that are missing (Don't know/No answer) were estimated using imputation techniques. In particular, a second scenario of simple imputation, in which we have imputed a previous vote to those respondents who gave their actual vote, but did not report vote recall, and a third setting of double imputation, in which we have attempted to impute a current and previous vote for respondents who do not answer either question.

Consequently, the problem now is choosing an imputation method from those that

TABLE VI. *Forecasts and estimation errors of the percentages of votes predicted using 2716 CIS poll data for the main parties contesting the 2007 Madrid Assembly elections*

Scenario(1)	Estimator	Percentages				Errors(3)			
		PP	PSOE	IU	OT	PP	PSOE	IU	OT
Oct-2003 Elections	Outcomes(2)	48.47	39.04	8.51	3.98	—	—	—	—
Raw Data n = 807	DIR	46.55	37.67	11.18	4.59	–6.58	4.34	2.32	–0.07
	PS	48.17	35.47	11.12	5.24	–4.97	2.14	2.26	0.57
	RAT	48.92	33.72	11.23	6.13	–4.22	0.39	2.37	1.46
	HD	50.23	32.93	11.94	4.91	–2.91	–0.41	3.08	0.24
Simple Imputation n = 881	DIR	47.65	36.42	10.67	5.26	–5.49	3.08	1.81	0.60
	PS	48.93	35.10	10.51	5.47	–4.21	1.76	1.65	0.80
	RAT	49.80	33.87	10.85	5.48	–3.34	0.53	1.99	0.82
	HD	50.68	32.91	10.99	5.42	–2.45	–0.42	2.12	0.75
Double Imputation n = 929	DIR	48.51	35.41	10.81	5.27	–4.62	2.07	1.95	0.60
	PS	49.57	34.33	10.46	5.64	–3.57	1.00	1.60	0.97
	RAT	50.19	33.38	10.64	5.78	–2.94	0.05	1.78	1.11
	HD	51.12	32.39	11.08	5.40	–2.01	–0.94	2.22	0.74
2007 Eleccions	Outcomesb	**53.14**	**33.33**	**8.86**	**4.67**	—	—	—	—

Source: Own elaboration using data from 2716 CIS survey.

(1) Raw data: Without imputation, only individuals for which current vote and vote recall are observed are used; Simple Imputation: Vote recall is imputed for those respondents for which current vote is observed; Double Imputation: Either actual vote, vote recall or both variables are imputed when unobserved; n effective sample size used.

(2) Percentage of valid votes recorded in the resident population.

(3) Computed as the average percentage difference between estimated and real values.

Table VII. *Goodness-of-fit measures between actual and estimated joint vote distributions obtained using the 2716 CIS survey. 2007 Madrid Assembly regional elections* (1),(2)

	Raw Data				Simple Imputation				Double Imputation			
Estimador	ENT	RMSE	AME	RME	ENT	RMSE	AME	RME	ENT	RMSE	AME	RME
DIR	5.31	4.11	3.33	13.27	4.23	3.29	2.74	13.19	3.42	2.73	2.31	12.47
PS	3.66	2.94	2.48	13.38	3.06	2.46	2.11	12.25	2.45	2.07	1.78	12.14
RAT	2.70	2.53	2.11	16.78	2.20	2.00	1.67	11.96	1.81	1.81	1.47	12.41
HD	1.99	2.13	1.66	11.65	1.69	1.68	1.44	11.51	1.63	1.61	1.48	11.87

Source: Own elaboration.

(1) Raw data: Without imputation, only individuals for which current vote and vote recall are observed are used; Simple Imputation: Vote recall is imputed for those respondents for which current vote is observed; Double Imputation: Either actual vote, vote recall or both variables are imputed when unobserved; n effective sample size used.

(2) ENT: Entropy; RMSE: Root of MSE; AME: Absolute mean error; RME: Relative average error.

have been proposed in the literature —regression imputation, random imputation, mean imputation, nearest neighbor imputation, multiple imputation, expert imputation, hot deck, or cold-deck (eg, Schafer , 1997; Särndal and Lundström, 2005, Galvan and Medina, 2007). In this case, in order to make the solution workable and meaningful, imputation by expert judgment was chosen (Särndal and Lundström, 2005: 164-5), using as informative variables interviewees' responses to questions such as: leaders evaluation, ideological self-identification, proximity to parties, ideological allocation of parties on behalf of the respondent, respondent's vote in other elections, evaluation of electoral outcomes and assessment of policies.

Tables VI and VIII display the results of the predictions obtained for the 2007 Madrid Assembly and the 2007 Barcelona local elections, respectively, after employing the three sets of data (without imputation, with simple imputation and double imputation) described in the paragraph above. Tables VII and IX, on the other hand, show the goodness-of-fit statistics of the estimated distributions.

As can be easily deduced by observing Table VI, the 2716 survey raw results are extremely biased and although imputations reduce the bias significantly, it remains significantly high (see DIR forecasts). In the case of Madrid, the predictions that best fit the actual outcomes are obtained using the HD estimator, displaying a significant advantage over RAT estimates, which rank second, and PS predictions, which are far from the actual results despite improving clearly on the direct forecasts (see Table VII). The combination of imputation and vote-recall correction seems to reduce bias and improves the overall accuracy of all the forecasts. However, this improvement does not affect all strategies equally. The HD estimator improves the least, while the other predictors further improve their performance after imputation with marked advances in their overall fit.

In the case of Barcelona, the results are somewhat different. In fact, the data are even slightly more biased after imputation than they were before, despite the raw data already showing significant levels of bias. Furthermore, this case highlights the huge overcorrection that vote recall induces in CiU and PP, and also in PSC, estimates. The latter probably explains why, without imputation, all estimators based on vote-recall corrections perform worse this time than the direct estimator (see Table VIII).

Out of the three estimators based on vote-recall, the RAT predictor performs clearly worse than the PS and HD strategies. The differences between the last two, however,

TABLE VIII. *Forecasts and estimation errors of the percentages of votes predicted using 2720 CIS poll data for the main parties contesting the 2007 Barcelona local elections*

Scenario(1)	Estimator	Percentages						Errors(3)					
		PSC	CiU	PP	ICV	ERC	OT	PSC	CiU	PP	ICV	ERC	OT
2003 Elect.	Out-comes(2)	33.57	21.45	16.12	12.07	12.81	3.97	—	—	—	—	—	—
Raw data n = 728	DIR	31.41	26.99	11.21	11.38	8.66	10.36	1.45	1.54	–4.40	2.03	–0.19	–0.43
	PS	27.65	22.47	16.17	12.59	8.82	12.30	–2.31	–2.99	0.57	3.24	–0.03	1.51
	RAT	26.15	21.16	17.43	10.51	9.02	15.73	–3.81	–4.30	1.82	1.16	0.17	4.95
	HD	27.85	22.17	17.50	11.16	9.26	11.07	–2.11	–3.28	1.89	2.81	0.40	0.29
Simle Imputation n = 826	DIR	32.31	26.58	9.37	11.66	8.99	32.31	2.36	1.12	–6.24	2.32	0.14	0.31
	PS	28.01	22.24	15.50	12.61	8.90	12.73	–1.95	–3.21	–0.10	3.27	0.05	1.95
	RAT	26.03	20.77	17.81	10.39	8.95	16.04	–3.93	–4.68	2.21	1.04	0.10	5.26
	HD	27.39	22.78	17.41	11.91	9.18	11.33	–2.57	–2.68	1.80	2.57	0.33	0.55
Double Imputation n = 888	DIR	33.26	26.90	8.83	11.50	9.24	10.27	3.31	1.44	–6.78	2.15	0.39	–0.52
	PS	27.36	22.06	17.00	12.65	8.96	11.97	–2.59	–3.40	1.39	3.30	0.11	1.19
	RAT	24.88	20.10	19.97	9.87	8.91	16.28	–5.08	–5.36	4.37	0.52	0.05	5.49
	HD	27.76	23.18	17.91	11.55	9.28	10.31	–2.19	–2.28	2.30	2.20	0.43	–0.47
2007 Elect.	Out-comes(2)	29.96	25.46	15.61	9.35	8.85	10.77	—	—	—	—	—	—

Source: Own elaboration using data from the 2720 CIS survey.

(1) Raw data: Without imputation, only individuals for which current vote and vote recall are observed are used; Simple Imputation: Vote recall is imputed for those respondents for which current vote is observed; Double Imputation: Either actual vote, vote recall or both variables are imputed when unobserved; n effective sample size used.

(2) Percentage of valid votes recorded in the resident population.

(3) Computed as the average percentage difference between estimated and real values.

TABLE IX. *Goodness-of-fit measures between actual and estimated joint vote distributions obtained using the 2720 CIS survey. 2007 Barcelona local elections*(1), (2)

	Datos brutos				Imputación simple				Imputación doble			
Estimator	ENT	RMSE	AME	RME	ENT	RMSE	AME	RME	ENT	RMSE	AME	RME
DIR	1.79	2.17	1.67	11.15	2.27	2.92	2.08	13.56	2.77	3.27	2.43	15.39
PS	2.04	2.14	1.77	12.02	1.96	2.18	1.76	12.23	2.34	2.33	2.00	13.08
RAT	3.24	3.22	2.70	16.92	3.46	3.44	2.87	17.79	4.32	4.16	3.48	20.52
HD	2.12	2.12	1.80	11.57	2.08	2.00	1.75	11.15	1.91	1.85	1.65	10.64

Source: Own elaboration.

(1 Raw data: Without imputation, only individuals for which current vote and vote recall are observed are used; Simple Imputation: Vote recall is imputed for those respondents for which current vote is observed; Double Imputation: Either actual vote, vote recall or both variables are imputed when unobserved; n effective sample size used.

(2) ENT: Entropy; RMSE: Root of MSE; AME: Absolute mean error; RME: Relative mean error.

are not conclusive. In terms of entropy, the PS estimator records the best results (see Table IX), although its advantage over the HD estimator is rather scanty and may even be questioned. Indeed, on the one hand we observe that, with double imputation, the HD estimator improves on the PS estimator in terms of entropy while, on the other hand, with raw and simple imputed data the differences between both sets of estimates (PS and HD) are minimal or even favour the PS. Choosing the best therefore depends on the indicator used to measure the goodness of fit. In fact, in the estimates without imputation, each estimator records the best fits for two of the four indicators when, in the other scenarios, the PS estimator yields the best just once. Nevertheless, any alleged advantage of the PS estimator would be based in this case on that fact that its predictions for PP were more accurate.

Conclusions

When dealing with votes counted, the superpopulation models based on the congruence that the aggregate electoral results of consecutive elections display on small scales (polling boxes, electoral sections, voting stations) have helped to significantly improve the predictions obtained with biased samples. Parallel to this, nonresponse bias is seen to be a growing problem in polls, where biased samples are the norm. The aim of this research is to study the predictive power of these methods in a survey environment and to assess their performance against the estimators currently used by the industry. In addition to this, the study also seeks to ascertain whether the use of a sampling selection procedure (probably less costly in terms of both money and time) whereby a greater number of electors in a smaller number of census tracts are selected (AL sampling) could be implemented without impairing the quality of the estimates.

In order to answer these questions, we have performed a complex simulation exercise for two different elections (2007 Madrid Assembly and 2007 Barcelona local elections) generating an enormous amount of samples under different scenarios of voter behavior when interviewed. Furthermore, to complete the research, the strategy has also been tested using real poll data. The actual data (with and without imputation, to reduce partial response) collected in two post-election surveys (2716 and 2720) conducted by the CIS were also analyzed. Four different estimators have been used to generate predictions: a direct estimator (DIR), which translates poll raw responses into percentages, and three additional estimators that make use of vote recall responses to improve forecasts, namely the weighted ratio estimator (RAT), the post-stratification estimator (PS) and the HD superpopulation estimator, which, to elaborate its forecasts, regresses the vote-recall corrected estimates obtained in the sampled sections on the outcomes recorded in those same sections in the previous elections.

In light of the results, we have obtained a valuable set of findings with broad practical impact. Despite the different political and geographic areas considered in the two simulation exercises implemented and the different hypotheses considered concerning the asymmetric response rates of voters, the conclusions reached in the simulations for both elections are very similar. In particular, it could be stated that:

i) All estimators generate highly accurate predictions under ideal conditions (without nonresponse or response errors).

ii) Introducing vote recall as an auxiliary variable in the forecasting process improves estimate accuracy, with the PS estimator producing the closest fit to raw data.

iii) Of all three vote recall based estimators, the PS estimator yields the best predictions in ideal conditions.

iv) The PS estimator, however, suffers greatly in presence of nonresponse bias and becomes less accurate the greater the bias.

v) The HD estimator is clearly the best predictor when more realistic sampling circumstances are considered (when nonresponse bias and response error appear in the samples).

vi) The RAT estimator generates (on average) quite similar solutions to the HD estimator, albeit generally less accurate.

vii) The difference in relative accuracy between the HD estimator and the RAT and PS predictors increases when nonresponse bias grows.

viii) Despite the obvious superiority that the HD estimator shows in more realistic situations, there still seems to be room for more accurate strategies.

ix) In general, selecting a smaller number of sections (along with drawing a larger number of interviews per section) slightly worsens estimate accuracy, albeit unevenly for each estimator. The HD estimator suffers the least from this change in sampling strategy.

x) As a rule, the combined use of the HD estimator and an AL sampling design would increase the quality of estimates and reduce monetary costs.

The results obtained using the actual data collected in both the CIS 2716 and 2720 surveys points in the same direction as those achieved in the simulations. On the one hand, working with the 2716 poll data, the HD estimator clearly dominates its competitors, being able to appreciably reduce the enormous bias of the original data. On the other hand, we find that both the PS and HD estimators generate quite comparable predictions with the 2720 survey data, HD performance nevertheless still constituting an improvement on direct estimates, which become more biased after imputation.

In view of the findings obtained in this study, the recommendation is clear: the HD predictor should be placed ahead of the PS and RAT strategies. Such a decision would almost certainly result in an average improvement in prediction accuracy. Furthermore, the quality of the estimates would not be significantly altered if a sample design in which fewer sections and more electors by section were sampled was adopted, particularly if this were accompanied by the use of the HD estimator. Therefore, taking into account that adopting such a plan would be less expensive, the only impediment that could initially discourage the industry from adopting it would be that such a change would affect accuracy when estimating other issues that are also included in electoral polls (such as leader evaluation). This potential limitation, however, could be adequately overcome by selecting sections in a completely non random fashion. This approach would be perfectly acceptable within the HD strategy, as it does not require a random selection to be applied. Adopting the HD approach would therefore also have the advantages of using purposive sampling, which could, *a priori*, guarantee adequate representation in the survey of the whole socio-political spectrum.

Despite the preceding statements, the HD estimator is not a panacea. On the one hand, the fact that the HD predictor on average achieves the best forecasts with realistic samples does not guarantee that it will generate better predictions than its competitors—the RAT and PS estimators—for a particular sample. Likewise, on the other hand, as HD bias figures remind us, there is still room for improvement in this context. In this regard, it would be interesting to explore whether expressly considering the geospatial dimension of the data and/or a more efficient use of sample and population information could lead to more accurate predictions.

In addition to the issues outlined in the paragraph above, which lead to other suggestive avenues of research using alternative approaches, there are at least a couple of topics within the approach taken in this paper that should be addressed in the future: (i) Deciding the sampling design that best suits the HD estimator and (ii) determining the most appropriate strategy for computing the sampling variance of the HD estimator taking into account its features.

REFERENCES

Aybar, C. (1998): *Modelos de Superpoblación y Poblaciones Discretas Finitas*, Tesis Doctoral, Universitat de València: Valencia.

Bernardo, José M. (1997): "Probing Public Opinion: The State of Valencia Experience", en *Bayesian Case Studies 3*, eds. C. Gatsonis, J.S. Hodges, R.E. Kass, R. McCulloch, P. Rossi and N.D. Singpurwalla, New York: Springer-Verlag, pp. 3-21.

Bernardo, José M. and F. Javier Girón (1992): "Robust Sequential Prediction from Non-random Samples: the Election Night Forecasting Case", in *Bayesian Statistics 4*, eds. J.M. Bernardo, J.O. Berge, A.P. Dawid and A.F.M. Smith, Oxford: Oxford University Press, pp. 61-77.

Biemer, Paul, Ralph Folsom, Richard Kulka, Judith Lessler, Babu Shah and Michael Weeks (2003): "An Evaluation of Procedures and Operations Used by the Voter News Service for the 2000 Presidential Election", *Public Opinion Quarterly,* 67: 32-44.

Curtin, Richard, Stanley Presser y Eleanor Singer (2005): "Changes in Telephone Survey Nonresponse over the Past Quarter Century", *Public Opinion Quarterly,* 69: 87–98.

de Leeuw, Edith and Wim de Heer (2002): "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison", in *Survey Nonresponse*, eds. R.M. Groves, D.A. Dillman, J. L. Etinge and R.J.A. Little, New York: Wiley, pp. 41-54.

Galván, Marco and Fernando Medina (2007): *Imputación de Datos: Teoría y Práctica*, Santiago de Chile: United Nations Publications.

Greene, William H. (2003): *Econometric Analysis*, New Jersey: Prentice Hall.

Groves, Robert M. (1989): *Survey Errors and Survey Costs*, New York: John Wiley & Sons.

Groves, Robert M. (2006): "Nonresponse Rates and Nonresponse Bias in Household Surveys", *Public Opinion Quarterly*, 70: 646-675.

Groves, Robert M. and Mick P. Couper (1998): *Nonresponse in Household Interview Surveys*, New York: Wiley.

Groves, Robert M., Don A Dillman, John L Eltinge and Roderick J.A. Little (2002): *Survey Nonresponse*, New York: Wiley.

Groves, Robert M., Eleanor Singer, Amy D. Corning and Ashley Bowers, (1999): "A Laboratory Approach to Measuring the Joint Effects of Interview Length, Incentives, Differential Incentives, and Refusal Conversion on Survey Participation", *Journal of Official Statistics*, 15: 251-268.

Gschwend, Thomas, Ron Johnston and Charles Pattie (2003): "Split-Ticket Patterns in Mixed-Member Proportional Election Systems: Estimates and Analyses of their Spatial Variations at the German Federal Election, 1998", *British Journal of Political Science*, 33: 109-128.

Isaki, Cary T., Julie H. Tsay and Wayne A. Fuller (2004): "Weighting Sample Data Subject to Independent Controls", *Survey Methodology*, 30, 35-44.

Johnston, Ron and Charles Pattie (2000): "Ecological Inference and Entropy-Maximizing: An Alternative Estimation Procedure for Split-Ticket Voting", *Political Analysis*, 8: 333-345.

Kalton, Graham and Daniel Kasprzyk (1986): "The Treatment of Missing Data", *Survey Methodology*, 12: 1-6.

Keeter, Scoot, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser (2000): "Consequences of Reducing Nonresponse in a National Telephone Survey", *Public Opinion Quarterly,* 64: 125–48.

King, Gary (1997): *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*, Princeton, NJ, Princeton University Press.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve (2001): "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation", *American Political Science Review*, 95 49-69.

Konner, Joan (2003): "The Case for Caution. This System Is Dangerously Flawed", *Public Opinion Quarterly*, 67: 5-18.

Martin, Elizabeth A., Michael W. Traugott and Courtney Kennedy (2005): "A Review and Proposal for a New Measure of Poll Accuracy", *Public Opinion Quarterly*, 69: 342-369.

Mitofsky, Warren J. (2003): "Voter News Service after the Fall", *Public Opinion Quarterly*, 67: 45-58.

Mitofsky, Warren J. and Edelman Murray (2002): "Election Night Estimation", *Journal of Official Statistics*, 18: 165-179.

Merkle, Daniel M. and Edelman Murray (2002): "Non-response in Exit Polls: A Comprehensive Analysis", in *Survey Nonresponse*, R. M. Groves, D. A. Dillman, J. L. Etinge and R. J. A. Little (eds.), New York: Wiley.

Pavía, Jose M. (2010): "Improving Predictive Accuracy of Exit Polls", *International Journal of Forecasting*, 26: 68-81.

—, Bernardi Cabrer and Ramón Sala (2009): "Updating Input-Output Matrices: Assessing Alternatives through Simulation", *Journal of Statistical Computation and Simulation*, 79: 1467-1498.

—, Beatriz Larraz, and Jose María Montero (2008): "Election Forecasts Using Spatio-Temporal Models", *Journal of the American Statistical Association*, 103: 1050-1059.

— -and López-Quilez, Antonio (2012): "Spatial Vote Redistribution in Redrawn Polling Units", under review.

Pavía-Miralles, Jose Manuel (2005): "Forecasts from Non-Random Samples: The Election Night Case", *Journal of the American Statistical Association*, 100: 1113-1122.

Pavía-Miralles, Jose Manuel and Beatriz Larraz-Iribas (2008): "Quick Counts from Non-Selected Polling Stations", *Journal of Applied Statistics*, 35: 383-405.

Randall, J. Jones, Jr. (2008): "The State of Presidential Election Forecasting: The 2004 Experience", *International Journal of Forecasting*, 24: 310-321.

Rodríguez Osuna, Jacinto (1991): *Métodos de Muestreo.* Madrid: Centro de Investigaciones Sociológicas.

— (2005): *Métodos de Muestreo. Casos Prácticos.* Madrid: Centro de Investigaciones Sociológicas.

Rosén, Bengt (1997a): "Asympotic Theory for Order Sampling", *Journal of Statistical Planning and Inference*, 62: 135-158.

Rosén, Bengt (1997b): "On Sampling with Probability Proportional to Size", *Journal of Statistical Planning and Inference*, 62: 159-191.

Särndal, Carl-Erik (2007): "The Calibration Approach in Survey Theory and Practice", *Survey Methodology*, 33: 99-119.

Särndal, Carl-Erik and Sixten Lundström (2005): *Estimation in Surveys with Nonresponse*, Chichester, England: John Wiley & Sons.

Särndal, Carl-Erik, Bengt Swenson and Jan Wretman (2003): *Model Assisted Survey Sampling*, New York: Springer.

Schafer, J.L. (1997): *Analysis of Incomplete Multivariate Data*, Boca Ratón, Florida: Chapman & Hall/CRC.

Singer, Eleanor (2006): "Introduction: Nonresponse Bias in Household Surveys", *Public Opinion Quarterly*, 70: 637-645.

Smith, Tom W. (1999): "Developing Nonresponse Standards", paper presented to the *International Conference on Survey Nonresponse*, Portland, October, 1999. http://cloud9.norc.uchicago.edu/dlib/nonre.htm.

Valliant, Richard, Alan H. Dorfman and Richard M. Royall (2000): *Finite Population Sampling and Inference. A Prediction Approach*, New York: John Wiley & Sons.

RECEPTION: 27/04/2010
ACCEPTANCE: 03/06/2011

Appendix

Appendix I: Estimation of cross-voting distributions (vote recall)

In order to estimate the cross-distribution of votes in each census section, a two-step procedure was carried out. In a first step, the voting transfer matrices obtained from the 2716 and 2720 poll data were adjusted to make them consistent with the actual results registered in each constituency. In a second step, a transfer matrix was estimated in each section by balancing the outcomes of the section to the matrix obtained for the whole district.[30]

By way of illustration, let us consider two successive elections with only 3 electoral choices (A, B or C) in which 1 million votes were cast. [We assume the same electors in both elections to simplify]. Let us say that in the first election each option, A, B and C, gained 500,000, 400,000 and 100,000 votes respectively, and let us assume that in the next election each party received 450,000, 440,000 and 110,000 ballots. Admit that a survey of 1,000 voters, in which each respondent was asked for his/her vote in the current and previous elections, were collected and that the results were classified in a contingency table as below, where the rows represent vote recalls and the columns the current election votes (for example, 55 voters said they had chosen option B in the previous election and option A in the current election).

	A	B	C	
A	400	80	10	490
B	55	350	15	420
C	3	10	77	90
	460	440	100	1,000

It is clear that this sampling transfer matrix is not completely consistent with the aggregate outcomes recorded in both elections (for example, according to the sample, party C would have obtained 90,000 votes in the first election, when in fact they gained 100,000). Nevertheless, taking the cross-distribution derived from the survey as a starting point, a voting transfer matrix among options and between elections could be approximated by imposing as constraints that row and column sums should match the actual values recorded. More specifically, using the RAS method (see, e.g., Pavia *et al.*, 2009) at the constituency level yields the following transfer matrix:

	A	B	C	
A	399,517	89,735	10,748	500,000
B	47,393	338,698	13,909	400,000
C	3,090	11,567	85,343	100,000
	450,000	440,000	110,000	1,000,000

After computing the estimated distribution of cross-voting at constituency level the same methodology is applied at section level (taking the cross-distribution estimated for the whole district as a reference). Thus, let us assume that in a district of 1,000 electors, 700, 250 and 50 voted for options A, B and C respectively in the previous elections and that in the current elections the outcomes were 630, 309 and 61. Then, applying the RAS matrix balance algorithm to these data, using the distribution estimated in the previous step as an initial approximation, we would

[30] From a practical-heuristic perspective, it is inefficient to work with all possible kinds of electoral behaviours. Consequently, in order to make the problem more manageable, the following were considered as possible categories to vote for: PP, PSOE, IU, Others or Blank, Abstention or No vote by age (this last option only for vote recall) in the case of the Madrid Assembly elections and, in the case of the Barcelona local elections, PSC, CiU, PP, ERC, ICV, Others or Blank, Abstention or No vote by age (the latter option only for vote recall.)

obtain the following cross-voting distribution for this section:

	A	B	C	
A	590	99	11	700
B	38	204	8	250
C	2	6	42	50
	630	309	61	1,000

By estimating these distributions for each census tract we can assign an actual vote and a vote recall to each elector after being drawn by choosing a section at random and within the section the elector also at random.

Appendix II: Selection of electoral sections and simulation of responses

Details concerning (i) the methods used to draw sections, (ii) the number of sections and electors selected and (iii) how survey responses were simulated are provided below. For CIS sampling, 101 sections were drawn in the region of Madrid, around 10 voters being chosen in each selected section (between 8-12). Meanwhile, in the city of Barcelona 10 voters were chosen in each of the 100 sections selected. Sections were selected following the stratified multistage cluster design typical of the polls conducted by the CIS (see Rodríguez Osuna, 2005). In the AL design the same strategy was employed in both elections. Twenty-five sections were randomly drawn (with selection probabilities proportional to the size of the section; Rosén, 1997a, 1997b) and 40 voters interviewed in each chosen section.

Once the sections that make up each sample were selected, we tackled the last stage: simulating electors' responses. To this end, we followed a hierarchical scheme of complexity regarding voter behavior when interviewed. Starting from the ideal case in which all voters contacted agreed to answer truthfully, we heightened the realism of the simulated samples to create situations where the two types of nonsampling errors with the greatest impact on the accuracy of forecasts (which are also the most common in survey studies) were recreated through voters' behavior (Groves, 1989; Särndal and Lundström, 2005, Pavia, 2010): Nonresponse bias and response error. In particular, we simulated samples of three types: (i) samples without error (WE), where each respondent faithfully reports his/her vote in the current and previous elections[31]; (ii) samples with nonresponse bias (NRB), where there are skewed distributions of nonresponse for the voters of each party, i.e., situations where the probability of an elector contacted providing an answer depends on his/her vote and differs according to it, and (iii) samples with nonresponse bias and response (measurement) error (RE), where in addition to the behavior described in (ii), a portion of the respondents report incorrect answers.

Electors were in all cases chosen by simple random sampling without replacement. In NRW and RE samples, nevertheless, each subject initially drawn was subjected to a dichotomous or Bernoulli trial[32] to decide whether or not the elector would become part of the final sample. Moreover, the interviewees finally selected in case (iii) were subjected to a second Bernoulli test to decide whether or not they would report truthful answers.

In case (i), in each selected section, m_s electors were randomly selected (with m_s approximately 10 for CIS sampling or equal to 40 in the AL design) and their votes in the past and current elections gathered. In cases

31 Although there are voters who do not vote and, therefore, cannot report any "casting vote", in order to simplify the language we include the option not to vote when referring to an elector vote.

32 Bernoulli trials are used to solve dichotomous decision problems, where the toss of a coin (with probability of heads and tails not necessarily being equal) is used to decide (as the result of the release) whether or not the selected subject performs a previously specified action.

(ii) and (iii), the voters of the section were first ordered and, depending on that order, a Bernoulli trial was performed to decide whether or not the voter would become part of the sample. This process was applied sequentially until either a sample of size m_s was obtained in the corresponding section or until all voters in the section had been subjected to the Bernoulli test.

The probability of each subject passing the first Bernoulli trial (and therefore becoming part of the sample) was not the same for all electors, this depending on their electoral behavior. In particular, the odds of belonging to the sample assigned to each subject were derived by comparing the results recorded in the corresponding elections and the estimates obtained directly from the baseline polls (2716 and 2720 CIS surveys). The probabilities were chosen to generate samples that replicated a nonresponse bias similar to that actually observed in the surveys used as models. In the case of the Madrid Assembly elections, it was assumed that among those electors who did not vote (in the current election) the percentage that would decline to take part in the survey would be between 50% and 60% (depending on the section), from where it was estimated that these percentages would fall between 15% and 25% for people who vote, of them about 70% (65%-75%) would be PP voters, around 22% (20%-25%) PSOE voters, approximately 5% (3%-7%) IU voters and about 3% (0%-5%) other option voters (OT). In the case of Barcelona, on the other hand, in each section the percentage of non-voters who would decline to be interviewed was assumed to be between 65% and 55%, and for voters they were: between 10% and 20% for PSC voters, 25%-15% for CiU voters, 60%-50% for PP voters, 23%-17% for ERC voters, 8%-2% for ICV voters, 40%-50% for other voters and 3% for blank voters. Applying these percentages to the recorded outcomes helps to simulate samples with similar biases to those observed in the baseline surveys, with the ranges of variation allowing certain flexibility in the structure of nonresponse in each section within a general pattern for the whole constituency.

Finally, in type (iii) samples the selected voters were additionally subjected to a second Bernoulli trial to decide whether or not the elector would report their actual vote[33]. Recall and current votes were randomly generated for the voters who passed this second Bernoulli test.

[33] The probability of an interviewee not reporting true responses was set *ad-hoc* at 5%.