

The Effect of Weighting and Multiple Imputation on Bias in Spanish Election Polls

El efecto de la ponderación y la imputación en el sesgo de los estudios electorales en España

Pablo Cabrera-Álvarez and Modesto Escobar

Key words

- Polls
- Voting Estimates
- Multiple Imputation
- Past Vote
- Weighting
- Total Survey Error

Palabras clave

- Encuestas
- Estimación de voto
- Imputación múltiple
- Recuerdo de voto
- Ponderación
- *Total Survey Error*

Abstract

The purpose of this article is to assess the effectiveness of post-survey adjustments made to electoral polls in order to correct for non-response bias. To do so we have used different weighting and multiple imputation methods using pre-election and post-election polls conducted by Spain's Centre for Sociological Research for all Spanish general elections since 1982. The results show the benefit of weighting by past vote when voters' preferences remain stable. However, the use of multiple imputation techniques to address missing values has a limited effect and is influenced by the variables included in the model.

Resumen

Este artículo tiene como objetivo evaluar la eficacia de las correcciones realizadas en encuestas electorales para ajustar el efecto de la no respuesta. Para ello se ponen a prueba distintos métodos de ponderación e imputación múltiple en todos los estudios preelectorales y postelectorales de elecciones generales al Congreso de los Diputados español llevados a cabo por el Centro de Investigaciones Sociológicas desde 1982. Los resultados muestran la ventaja de utilizar la variable recuerdo de voto en la ponderación cuando hay estabilidad en las preferencias de los electores. De modo complementario, para tratar la no respuesta, el uso de técnicas de imputación tiene un efecto limitado y condicionado por las variables incluidas en el modelo.

Citation

Cabrera-Álvarez, Pablo and Escobar, Modesto (2019). "The Effect of Weighting and Multiple Imputation on Bias in Spanish Election Polls". *Revista Española de Investigaciones Sociológicas*, 165: 45-64. (<http://dx.doi.org/10.5477/cis/reis.165.45>)

Pablo Cabrera-Álvarez: Universidad de Salamanca | pablocal@usal.es

Modesto Escobar: Universidad de Salamanca | modesto@usal.es

INTRODUCTION¹

Estimates of voting based on pre-election polls tend to approximate final election results the closer they are carried out to the actual election. Although in recent years studies suggest that the accuracy of pre-election polls continues to meet satisfactory limits of what is expected (Jennings and Wlezien, 2018), we have also seen cases, such as the general elections in the United Kingdom in 2015 and in Spain in 2016, that have caused widespread debate about the usefulness and need for pre-election polling.

Among the reasons for the lack of accuracy in pre-election polls is the existence of non-response bias, that is, a systematic difference between the voting intentions of those who participate in the study and those who do not. To mitigate this bias, there are statistical techniques that correct possible deviations in the sample profile using complementary information. One of the commonly used auxiliary variables in this procedure is past vote, based on asking survey participants about their past voting behaviour. However, debate continues over the use of this variable, as there is no evidence that it always has a positive effect on the accuracy of voting estimates. With the same aim, multiple imputation (MI) techniques can be used to assign valid values to those who say they do not know their voting intention or do not answer the question.

This article, focused on the Spanish case, seeks to determine the effect of the use of multiple imputation and weighting on the accuracy of voting estimations, using a perspective that compares pre and post-electoral polls for Spanish parliamentary elections in the period from 1982 to 2016.

Compared to previous research (Escobar *et al.*, 2014; Pavía and Larraz, 2012; Rivas *et al.*, 2010), this study represents a new and important contribution for three reasons. The first is the use of different transformations of the past vote variable in the weighting, along with a set of socio-demographic variables that have not been previously used in Spain. Second is the extensive use of multiple imputation to address both voting intention and past voting. The third is that both the weights and imputations are tested using pre and post-electoral studies by Spain's Centre for Sociological Research (CIS), covering general elections in Spain since 1982. This perspective over time is necessary to clarify if the success of these techniques changes from one election to another and if there is a trend that explains such variability.

The article is divided into four sections. In the first, we look at how accuracy is defined and propose the *Total Survey Error* (TSE) paradigm as the theoretical framework for studying election polls. Next, we present our working hypotheses and the data and methodology used. This is followed by the results of our analysis and discussion. Lastly, we present our conclusions.

ACCURACY AND SOURCES OF ERROR IN PRE-ELECTION POLLS

The concept of accuracy in election polls involves two issues. On the one hand, the notion of variability resulting from sample size and variance in estimation, represented by the margin of sampling error, and on the other, the notion of fit, as the difference between the estimation and population data, in this case, the results of the election. In studies such as this one, which analyse poll performance, the concept of fit is used taking into account that this comparison between polls and election results can be altered by the effects of the electoral campaign (Crespi, 1988; Sturgis *et al.*, 2016).

¹ This research has been supported by the predoctoral scholarship programme of the Obra Social "La Caixa".

The lack of accuracy in pre-election polls is a recurring theme in the literature on public opinion and elections (Caballé *et al.*, 2013; Callegaro and Gasperoni, 2008; Durand *et al.*, 2004; Sanders, 2003; Traugott, 2005). Among the causes of the discrepancy between opinion polls and actual election results, we find changes in the preferences of voters between the time of the survey and the elections (Abrams, 1970; Shlapentokh, 1994), the methods used to determine the likelihood of voting (Durand *et al.*, 2004; McEwen, 2004; Sturgis *et al.*, 2016), the sample used (Abramson, 2007; Curtice, 1997; Lynn and Jowell, 1996; Worcester, 1996), problems related to population coverage (Callegaro and Gasperoni, 2008; Durand *et al.*, 2001; Sauger, 2008), and how cases that do not respond to the survey or to the question regarding voting intention are treated (Anderson, 1992; Jowell *et al.*, 1993; Katz, 1941).

The TSE framework allows us to systematically analyse the sources of error that exist in the design of the study, as well as in the gathering, processing and analysis of data. Knowing and controlling these sources of error is essential for maintaining the quality of estimates (Biemer, 2010; Biemer and Lyberg, 2003). The TSE divides sources of error into two groups: those related to measurement and those related to representativeness. Regarding measurement, we find issues of validity, measurement errors and errors in data processing. Regarding representativeness, we find errors in coverage, sampling error, non-response error and errors derived from the adjustments made after the gathering of the data (Groves *et al.*, 2013).

Non-response bias in pre-election surveys

This study focuses on the non-response error. Non-response refers to a lack of information due to a sample element not being

reached or not collaborating in the survey or a part of the survey (Lynn, 2008). In the context of this study, we say that there is a non-response bias when those that respond in a poll, or specifically to a question regarding voting intention, have voting preferences that are different from those that do not respond.

Unit non-response (that is, when the sample element rejects participation in the survey or is not contacted) has been identified as one of the causes of problems in the accuracy of pre-election polls in different countries, among them Spain (Durand *et al.*, 2004; Jowell *et al.*, 1993; Smith, 1996; Pavía *et al.*, 2016). In this regard, some studies have shown that the propensity to respond to socio-political surveys is related to the level of interest in politics the citizens in the sample have (Voogt and Saris, 2003). More recently, it has been shown that those who are going to vote tend to be over-represented in post-election studies, contributing to estimates of participation based on polls exceeding real figures (Ansolabehere and Hersh, 2012; Sciarini and Goldberg, 2016).

The other factor associated with non-response bias is the refusal of some participants to reveal their voting intention. This has also been identified as a possible cause of the lack of accuracy in pre-election surveys (Curtice, 1997; McEwen, 2004; Sauger, 2008). Regarding Spain, Urquizu (2005) showed that in the 1980s conservative voters were less likely to reveal their voting preferences, while this tendency was reversed in the 1990s.

Weighting and imputation as methods for adjusting non-response bias in studies of voting

Once the data have been gathered, it is possible to apply adjustments to reduce the impact of biases caused by both unit and item non-response. In the case of total

non-response, weighting techniques are used to rebalance the final sample. In the same way, imputation techniques can be used to attribute valid response categories to those that avoid answering a specific question.

The use of weighting and similar techniques to calibrate results has been common in combination with quota sampling (Särndal, 2007). However, there is a debate in the literature over the usefulness of including past vote as an auxiliary variable in the weighting. For example, in the United States the majority of polling firms have avoided using this variable (Voss *et al.*, 1995). In a classic study on pre-election polling methodology in the U.S., Crespi (1988: 40-41) states that the main reason cited for not using this variable is measurement problems associated with it, including the over-representation of voters in comparison to abstentionists, and the over-representation of those who choose the winning candidate or party versus those that vote for the losers. Worcester (1996), in regard to the United Kingdom, argued that the use of past vote does not help and could actually lead to less accurate estimates. Along the same lines, a recent study on pre-election polls in Canada and France by Durand *et al.* (2015) shows that past vote can both improve and worsen the accuracy of voting estimates.

Despite this debate, the use of the past vote variable in weighting is common. For example, in the United Kingdom and in France this variable has been used to correct for under-representation of conservative voters (Crewe, 2001) and National Front voters (Durand, 2008) respectively. In Spain, only a few academic studies have attempted to shed light on this phenomenon. Escobar *et al.* (2014) compared different methodologies used to carry out estimates of voting behaviour based on polls using past vote as a weighting variable. In that study, they found that in elections that produced a change in the ruling party, the use of the past vote var-

iable weakened the accuracy of estimates in the period 1979 to 2011. Pavia and Larraz (2012) also experimented with different forms of weighting by past vote, reaching the conclusion that post-stratification was not the best method in response to non-response bias.

Regarding item non-response, once the data have been gathered, researchers have to decide how to proceed with cases that are likely to vote but that do not reveal their voting intention. This problem has been addressed using *ad hoc* techniques defined by each research organisation (Crespi, 1988; Lynn and Jowell, 1996; Sturgis *et al.*, 2016). In Spain, we find the study of Varela *et al.* (1998), which describes different methods that can be used for imputing a valid response to those who do not reveal their voting preferences, as well as that of Pavia and Larraz (2012), who employed imputation by expert criteria to address non-responses in voting intention and past vote. In addition, Rivas *et al.* (2010) study discusses the relevance of the use of imputation to address partial non-response in voting intention, using the 2000 elections as a case study. Their conclusion is that this technique is only effective when the predictors allow us to differentiate all the possible categories of voter intention.

Multiple imputation is a technique that is used to assign valid values to cases that have missing values, but its application in pre-election studies has been limited. King *et al.* (2001) analysed its potential use in political science, suggesting it could be used to study the preferences of non-voters. Bernhagen and Marsh (2007) used this technique to assign valid values to individuals that did not declare their voting intention, and Liu (2014) used multiple imputation techniques to assign preferences to individuals that did not reveal their voting intention in a pre-election study in Taiwan, but without success. For Spain, Escobar and Jaime (2013) were also not able to achieve greater

accuracy in their estimates using different methods of multiple imputation in pre and post-election studies from the Centre for Sociological Research in the 2011 general elections.

RESEARCH HYPOTHESES

To meet our objective of determining the effect of the use of weighting and multiple imputation on the accuracy of estimates of voting behaviour in Spain, we have formulated the following hypotheses:

Hypothesis 1: Weighting the sample using socio-demographic variables positively affects the accuracy of vote estimates.

However, we expect the impact to be limited because quotas are already used and because in general these types of variables have little relationship to voting intention. In addition, weighting techniques reduce the total non-response bias in estimates when the information employed is correlated with the propensity to respond to the variable of interest, in this case, voting intention (Särndal, 2007).

Hypothesis 2: Weighting the sample by past vote increases the accuracy of vote estimates.

Using the variable past voting behaviour, quite widespread in the polling industry, has a different impact, given that it is correlated with intention to vote (Crespi, 1988; Crewe, 1997). Some studies in other countries, however, have shown that the effect of using this variable has been minimal (Duran, Des Lauriers and Valois, 2015).

Hypothesis 3: The use of multiple imputation techniques in treating the past vote variable reduces the level of bias present in this variable, and as a consequence, the use of this imputed variable in the weighting increases the accuracy of the vote estimates.

To the extent that the recall of past vote in previous elections could be impacted by memory problems or by item non-response bias (Crespi, 1988; Worcester, 1996), we propose a procedure to correct deviations by the use of multiple imputation techniques.

Hypothesis 4: The use of weighting that combines socio-demographic variables and past vote will be the most effective in reducing the error level of vote estimates (PV+SD).

This hypothesis is a corollary of the preceding ones. If weighting with socio-demographic variables and past vote separately affects vote estimates positively, it would be expected that combining these variables would improve the results.

Hypothesis 5: The use of the past vote variable to weight results has a positive effect on the accuracy of estimates in elections in which there is political continuity.

Escobar *et al.* (2014) reveal a trend in the use of past vote in weighting in the Spanish case (1979–2011): when in certain elections the ruling government fails to maintain its hegemony, the use of past vote in weighting has a negative effect on accuracy. In the 2014 elections, Spain's party system changed with the emergence of two new parties and the decline in the percentage of support given to the two main political parties (Orriols and Cordero, 2016; Rama, 2016). It is necessary to see whether this generalisation is confirmed in the subsequent 2015 and 2016 elections.

Hypothesis 6: The use of multiple imputation techniques to assign a voting intention or behaviour to those that do not know or do not answer increases the accuracy of vote estimates.

In this regard, we advocate for the use of multiple imputation to assign valid values to those that do not reveal their voting behaviour in elections being held close to the date of the survey (King, 2001).

Hypothesis 7: In the same way that occurs with weighting by past vote, the effect of the use of multiple imputation techniques depends on whether election results return the ruling party to government or not.

Escobar and Jaime (2013) showed the positive effect that imputation had on the accuracy of estimates in the 2011 elections. However, to date, no other study on the effects of this procedure on other elections has been published.

Sample

In order to examine the implications of using different types of weightings and imputations, we have worked with pre-election and post-election polls carried out by the CIS between 1982² and 2016 for Spain's general elections. All of the CIS samples used multi-stage stratified sampling by province and population size, with the selection of households by random routes and subjects by sex and age quotas. The studies used are shown in the following table:

TABLE 1. Year, number and size of CIS studies used

Year	CIS study number	Pre-electoral sample size	Post-electoral sample size
1982	1.326 y 1.327	24,832	2,394
1986	1.526 y 1.542	25,304	6,842
1989	1.821/37 y 1.842	27,122	2,508
1993	2.060 y 2.061	2,462	4,225
1996	2.207 y 2.210	6,544	4,610
2000	2.382 y 2.384	24,040	4,386
2004*	2.555	24,109	
2008	2.750 y 2.757	18,221	5,247
2011	2.915 y 2.920	17,201	6,056
2015	3.117 y 3.126	17,403	5,457
2016	3.141 y 3.146	17,458	5,136

* The 2004 post-electoral study was excluded from the analysis as it did not include the variable, past vote.

Note: Starting in 2000, with the exception of 2004 and 2016, the pre and post-electoral studies were panel type studies.

METHODOLOGY

In this part, we present the methodology in four sections. In the first one we describe our data sources. In the next section, we address the different criteria used for weighting the data, and in the third section, the imputation procedures used. In the last section we present the criteria used to determine the accuracy of voting estimates.

The inclusion of post-electoral studies stems from the limitation resulting from the dates of the fieldwork for the CIS pre-election polls, around one month before the elections. Post-election studies seek to mitigate the

² The 1977 and 1979 elections are not included because no post-electoral study is available. In addition, in the case of the 1977 elections, no past vote variable, essential for this study, is available.

possible bias introduced by the effects of the electoral campaign, which are not detected in the pre-electoral polls. However, it is also necessary to point out limitations presented by post-electoral studies: 1) voters in elections are over-represented (Sciarini and Goldberg, 2016); 2) in the case of panel-type studies, these may have conditioning effects on participants (Sturgis *et al.*, 2009), and 3) there is usually an over-representation of the winning party in the recent elections (Crespi, 1988).

Weighting criteria

In pre-electoral studies, the CIS designs a stratified sample by province with non-proportionate allocation. In these cases, a selection weight has to be applied. These weights (w_k) are equal for interviewees in the same electoral district, and their formula is the following:

$$w_k = e_k / n_k$$

with e_k being the size of the electoral census and n_k being the number of interviews ca-

ried out in each strata, province or electoral district.

The remaining weights were calculated using the logistic calibration method. There are other methods for generating weighting coefficients, such as the use of non-response models to determine the probability of responding in a survey, or methods based on *propensity score matching* techniques. Comparison of these methods shows that the key lies in the predictors used (Mercer *et al.*, 2018), rather than in the statistical technique used to generate the weights. In this case, given population data can only be obtained in aggregate form, the technique used was calibration in its logistic version; its advantage over the linear version is that it avoids the generation of negative weighting coefficients.

After obtaining the sample, logistic calibration requires a comparison of the distribution of one or more of its variables to see if they coincide with the parameters of the population, in order to calculate weights that ensure that the sample results coincide with

TABLE 2. Weightings by province, past vote (PV) and sociodemographic variables used in the design of the research

Abbreviation	Weighted variables	Imputed PV	PV sphere *
BE	Province		
SD	Province and sociodemographics**		
PV0	Province and past vote (PV)	No	NR, NV and NVM excluded
PV1	Province and past vote	No	NR, NV and NVM included
PV2	Province and past vote	Yes (NR imputed)	NR, NV and NVM included
PV3	Province and past vote	Yes (NR and NVM imputed)	NR, NV and NVM included
SD+PV	Province and sociodemographics** and PV	Yes (NR and NVM imputed)	NR, NV and NVM included

NR: non-response (DK (don't know) and NA (no answer)); NV: did not vote in prior elections; NVM: did not vote, minor.

* Non imputed categories, whether for no response (NR) or for not voting due to being a minor (NVM), only the weight corresponding to the rest of the weighting criteria are assigned.

** Age by sex, autonomous region, size of habitat, education level and employment status.

the population totals in the selected variables³. The first calibration criterion used includes only socio-demographic variables, while the rest of the weights use past vote⁴. A summary of the weights used is provided in Table 2.

Imputation criteria

The method proposed by Rubin (1987) was used for the analysis and imputation of incomplete data, which consists of reconstructing new data sets, as many as the researcher establishes, with randomly simulated values based on other variables from the study that contain more complete information⁵. In contrast to a single imputation, which consists of estimating the data only once, multiple imputation makes a series of estimates – by simulating a number of complete data sets – from which a single estimate can be constructed, supplemented by the variation from the diverse estimates carried out. Consequently, the variances of the parameters can be obtained more accurately than with a single imputation.

³ For more information regarding the calculation of weighting coefficients by calibration, see the studies of Särndal (2005) and Lundström and Särndal (2001). The calibration was carried out in Stata, using the calibrate package designed by D'Souza (2011).

⁴ To carry out the calibration it is necessary to have the population distributions for the auxiliary variables. The data to carry out the weighting by past vote and population size come from Spain's Interior Ministry. The information on population distributions for the variables sex, age and autonomous regions come from Spain's National Statistics Institute (INE). Historical data on education level and employment status come from studies by Fuente and Domenech (2015) and Fuente (2015), respectively. In the case of the last two elections, 2015 and 2016, for which no data exists in the just mentioned studies, values for the populations were obtained from the INE's Labour Force Survey for employment status and were interpolated for education level.

⁵ A basic introduction along with the way to obtain these models with Stata can be found in the book dedicated to multiple imputation (Stata, 2015). In addition, a theoretical and applied presentation in Spanish is found in the already cited book by Rivero (2011).

There are different imputation procedures based on Bayesian and frequentist principles. Essentially, we can distinguish between univariate imputations (one variable at a time), based on the posterior predictive distribution of missing data, and chained imputations, which involve feedback of the imputed variables. Table 3 summarises the types of imputations taken into account in this study (note the correspondence between the versions of imputed past vote and the weights of the same name).

The choice of the predictors included in the imputation models was made based on theoretical criteria (Escobar and Jaime, 2013) and considering the limitations resulting from the research design. First, studies of electoral behaviour were used to determine which socio-demographic and political predictors are related to voting intention (for example, Bosch and Riba, 2005; Jaime and Saéz, 2001; Lago and Lago, 2005). Secondly, the list was limited to be able to apply the same model to all the elections studied.

Given that the variable of interest for the imputation is always the vote (x_i), we used a multinomial model in which the probabilities for the k categories of the variable respond to the following formula, where \mathbf{z}_i is the vector of the variables used in the imputation:

$$\Pr(x_i = k | \mathbf{z}_i) = \frac{1}{1 + \sum_{l=2}^K \exp(\mathbf{z}_i' \boldsymbol{\beta}_l)}, \text{ si } k = 1$$

$$\Pr(x_i = k | \mathbf{z}_i) = \frac{\exp(\mathbf{z}_i' \boldsymbol{\beta}_k)}{1 + \sum_{l=2}^K \exp(\mathbf{z}_i' \boldsymbol{\beta}_l)}, \text{ si } k > 1$$

Evaluation of the accuracy of results

To assess the results of the estimates and imputations we use the weighted mean absolute error (WMAE), used in the literature on

TABLE 3. *Imputations included in the design for past vote and voting intention*

A) Past vote (weighting for this variable)		
Name*	Objective variable (model)	Predictive variables
PV1 not imputed		
PV2 imputed (NR)	Past vote (multinomial)	Mixed set ** Mixed set
PV3 imputed (NR and NVM)	Past vote (multinomial)	Mixed set ** Mixed set
B) Voting intention (to estimate value)		
Name	Objective variable (model)	Predictive variables
1. Not imputed		
2. Basic univariate	Voting intention (multinomial)	Basic set ***
3. Enhanced univariate	Voting intention (multinomial)	Enhanced set ****
4. Basic chained	Voting intention (multinomial) Past vote (multinomial) Ideology (ordinal)	Basic set ***
5. Enhanced chained	Voting intention (multinomial) Past vote (multinomial) Ideology (ordinal)	Enhanced set ****

* In past vote, the non-response (NR) or those who did not vote in the previous elections because they were minors (NVM) were imputed (univariate method).

** Sex, age, habitat size, autonomous community, voting intention and ideology.

*** Sex, age, education level, past vote and ideology.

**** Sex, age, education level, past vote, ideology, evaluation of the economic situation and evaluation of the political situation.

forecasting time series⁶. The formula is the following:

$$\epsilon = \sum_{k=1}^K |\hat{p}_k - p_k| p_k$$

where p_k is the electoral result for each k party and \hat{p}_k their corresponding estimates.

⁶ There are three measures used for these ends: absolute mean error, quadratic mean error and standardised quadratic mean error. For their visibility, we have used the first, adding the weighting of the average errors. The latter was done because we are forecasting non-bipartisan elections, so it is logical to give greater importance to the errors committed in regard to the most voted parties. See Lewis (2005) and Hyndman and Koehler (2005).

Models

To discover which weighting and imputations methods are the best for election forecasting, we have considered two results:

- a) The WMAE(S) independently obtained for each survey based on voting intention (or past vote in the case of post-election studies) for the elections considered (models 1 to 4 in table 1 of the appendix).

The predictors included in these models were:

- 1) Year of the election (models 1 and 2).
- 2) Classification of elections as elections of change or continuity (models 3 and 4).
- 3) Type of survey (pre or post-election).

These two variables serve to control for the effect of the electoral campaign and the political climate.

- 4) Weighting modalities (the seven presented in table 2), with the aim of verifying *hypotheses 1, 2 and 4*.
- 5) Methods of imputing voting intentions (the five presented in table 3.b⁷), to verify *hypothesis 6*.
- 6) Interaction between the form of weighting and the year (models 1 and 2), or classification of the elections based on whether they are elections of change or continuity (models 3 and 4), considered in *hypothesis 5*.
- 7) Interaction of the imputation method and year (models 1 and 2), or classification of the elections based on whether they are elections of change or continuity (models 3 and 4), considered in *hypothesis 7*.

b) The WMAE(P) of past voting in elections prior to the poll in question (model 5 of table 1A in the appendix) for testing *hypothesis 3*. In this case, the predictors were year, type of survey and method of imputation of past vote (the three included in table 3.a).

With the different treatments of imputation and estimation by election and survey, we obtained 595 different estimates⁸ of the WMAE(S) and 63 of the WMAE(P).

⁷ But only three imputation modalities are possible before 2000, as the surveys did not include questions on evaluation of the government and the economy. As a result, models 1 and 3 are split (in which we only consider three imputation modalities for the whole period analysed) in models 2 and 4 in which there are more imputation modalities, but over less time: only the last six elections.

⁸ Among them, 441 do not contain broader imputations and, therefore, they are available in the 11 elections addressed, while 385 correspond to the surveys in which participants were asked to evaluate the government and the economy and, therefore, exclude predictions prior to the year 2000.

The hypotheses mentioned were compared within regression models, adjusted with least squares, through specific contrasts of the estimated means of the weighted mean absolute errors. Once the F values of these comparisons were calculated, the Bonferroni correction was applied to avoid Type I errors (Rosenthal *et al.*, 2000).

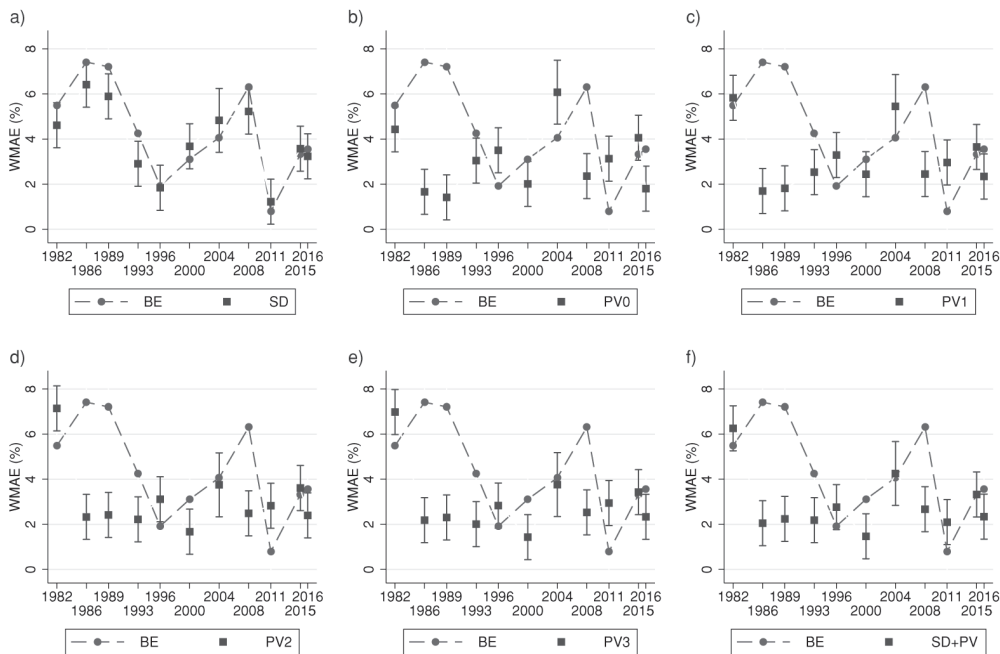
RESULTS AND DISCUSSION

The first hypothesis proposed in this study refers to the need to use socio-demographic weighting to improve the accuracy of voting estimates. We stated that the improvement in accuracy after balancing the sample by sex, age, education and economic activity would depend on whether these variables are related to voting intention and the probability of responding to the survey (Särndal, 2007). Graph 1.a shows that during the 1980s, the use of this weighting contributes slightly to improving estimates, although if the complete period is analysed there are no differences between the estimates without weighting and adjusted estimates by socio-demographic profiles ($F_{(1, 341)} = 2.59$; $p = 0.650$). The fact that socio-demographic variables do not have a clear relationship to voting intention or the likelihood of responding in a survey is not surprising, as other studies in the American and British contexts also point in this direction (Crespi, 1988; Sturgis *et al.*, 2016).

Faced with the ineffectiveness of using socio-demographic factors, an alternative is to use past voting, which is related to voting intention, and with the likelihood of accepting to participate in a survey (Voogt and Saris, 2003). The results show that in general the use of this variable in weighting helps reduce the error level in voting estimates (graphs 1.b to 1.f).

As described in the methodology section, we have worked with several versions of weighting by past vote in order to understand

GRAPH 1. Graph 1 Comparison of the Weighted Mean Absolute Error (WMAE) for estimating the weighted vote in its different versions in comparison with the base estimate (BE) without weighting



The reference for comparison is the base estimate (BE).
 SD: Sociodemographic (sex, age, region, employment status and qualifications); PV0: Past vote excluding those DK/NA or did not vote; PV1: Past vote;
 PV2: Imputed past vote (DK and NA); PV3: Imputed past vote (DK, NA and underage previous elections); SD+PV: Sociodemographic and PV3.

if the treatment of this variable through imputation has a positive impact on the quality of estimates (hypotheses 2, 3 and 4). Two of these alternatives (PV0 and PV1) do not use imputation techniques to correct for possible bias resulting from item non-response, while the other three versions do use this technique (PV2, PV3 and SD+PV).

In general, although public information regarding this is limited, we know that polling firms carry out a minimum transformation of the past vote variable before including it in any weighting. In this study, we have tried to replicate this strategy in two weightings. In the case of PV0, we excluded from the process those who do not remember or do not respond to the question regarding past vote, as well as those who said they did not vote in previous elections. In the majority of the elections studied, the use of this weighting improves the accuracy of vote estimations,

in comparison to the use of socio-demographic weighting ($F_{(1, 341)} = 15.98; p = 0.001$) or the absence of weighting ($F_{(1, 341)} = 31.44; p < 0.001$). The same behaviour is found using the PV1 weighting, which in contrast to the case of PV0, includes those respondents who do not reveal their votes in prior elections or who did not vote (without weighting: $F_{(1, 341)} = 27.22; p < 0.001$; socio-demographic: $F_{(1, 341)} = 13.01; p = 0.002$). This overall positive effect contrasts with the reticence shown by Worcester (1996) or the findings presented by Durand *et al.* (2015) for French and Canadian elections. Despite the deficiencies that this variable may present, overall its use is positive in the Spanish case for the period studied (hypothesis 2).

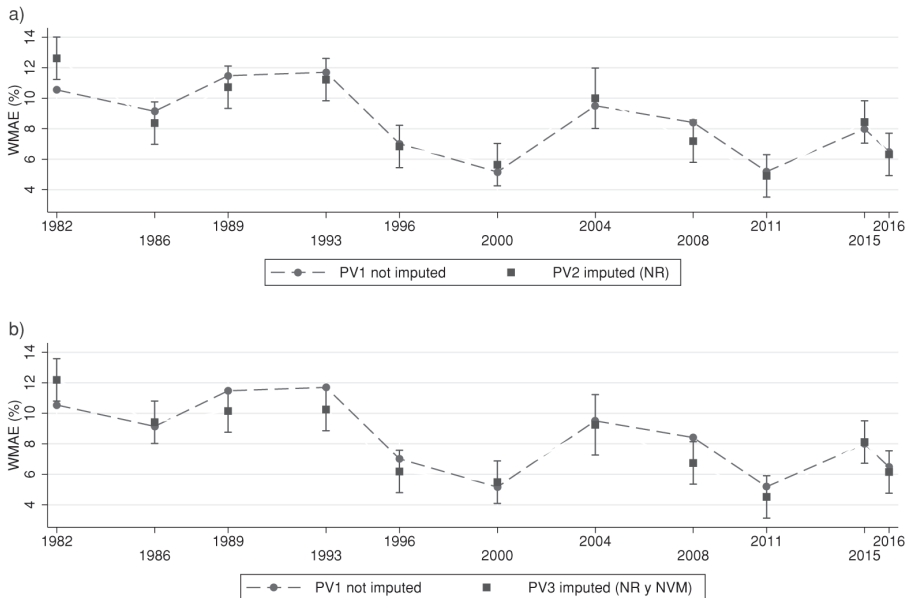
A possible improvement in the weighting of past vote in prior elections would consist of correcting for deviations in this variable, imputing valid values to those who did not

answer. The PV2, PV3 (hypothesis 3) and SD+PV (hypothesis 4) versions are variations in the weighting by past vote, in which this variable was previously treated with multiple imputation techniques. In the case of PV2, a valid value was assigned to those who did not respond, and with PV3, a valid value was also imputed to those who were not old enough to vote in prior elections. In both cases (graphs 1.c and 1.d), the use of the past vote variable increases the accuracy of estimates in comparison to the absence of weighting (PV2: $F_{(1, 341)} = 29.35; p < 0.001$; PV3: $F_{(1, 341)} = 34.95; p < 0.001$) or the use of socio-demographic weighting (PV2: $F_{(1, 341)} = 14.50; p = 0.001$; PV3: $F_{(1, 341)} = 18.51; p < 0.001$). However, there is no significant improvement if we compare the results with versions of the past vote without imputation (PV0 and PV1). In the case of weighting SD+PV, which combines socio-demographic variables with imputed past vote (PV3), we do not obtain better results than those found

with the rest of the weightings of past vote, presumably due to the null effect of socio-demographic variables in the majority of elections.

The use of imputation to reduce the bias caused by item non-response in the past vote variable does not improve the accuracy of estimates weighted by this variable. This can be seen in graph 2, as the corrections applied to assign valid values to those who did not respond are small (PV2 imputed (NR)), or to those who did not answer and to those who could not vote in prior elections (PV3 imputed (NR and ME)), barely reduce the error level of the variable with the exception of the 1993 and 2008 elections, where we find improvements of more than one percentage point. This may be due to two factors: the predictors included in the imputation and the magnitude of the bias due to partial non-response. On the one hand, the predictors included in the imputation models may not effectively discriminate among the

GRAPH 2. Comparison of the Weighted Mean Absolute Error (WMAE) of the variable past vote without imputation (PV1) and the variables imputed past vote (PV2 and PV3)



The reference for comparison is the base estimate (PV1 not imputed). PV2 imputed (NR): Imputed non-response (DK and NA); PV3 imputed (NR and NVM): Imputed non-response and underage in previous election (NVM).

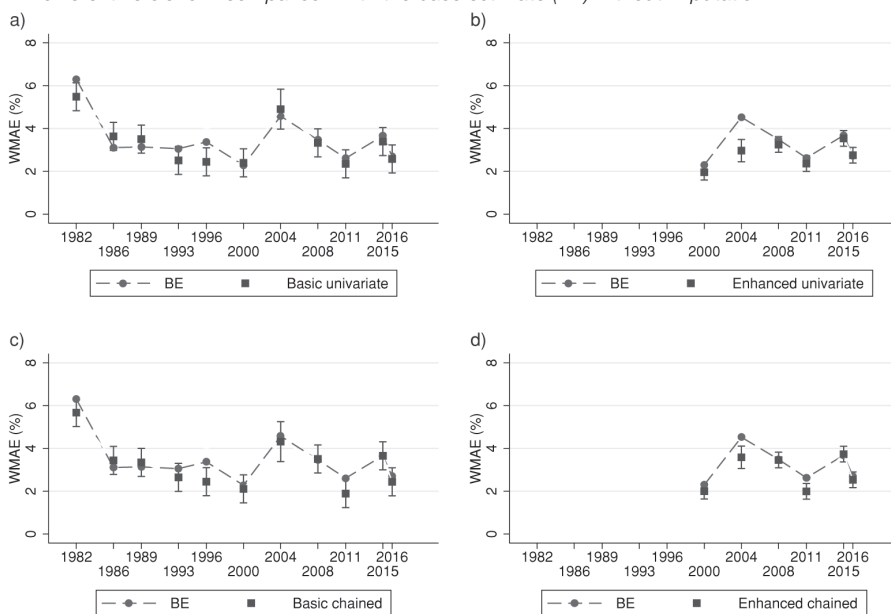
different options regarding past vote. On the other hand, the bias generated by partial non-response is likely to be minimal and despite its correction, there may still be a high level of error in past vote recall due to other factors, such as unit non-response. For example, the latter would occur if all those who respond to the survey correctly recall their voting behaviour in past elections, but their voting intentions are different from those who decided not to participate in the survey.

This positive use of past vote in weighting is however not uniform in the period studied (hypothesis 5). In line with the findings of Escobar *et al.* (2014), in the elections in which a change of government took place (1982, 1996, 2004 and 2011), the use of past vote has no positive effect on estimates (for example, PV3: $F_{(1, 338)} = 5.39, p = 0.251$). This trend continued in the new political cycle that began in 2014, characterised by the emergence of Podemos and Ciudadanos in the party system (Orriols and Cordero, 2016;

Rama, 2016), so that in the 2015 elections, the use of past vote does not contribute to improving the estimate (for example, PV3: $F_{(1, 404)} = 0.01, p = 0.999$), while it does seem to do so in surveys of the 2016 elections, although the comparison test does not provide robust results (for example, PV3: $F_{(1, 404)} = 1.66, p = 0.198$).

In estimating the vote, it is also important to look at the possible deviations from voting intention, due to the fact that those who reveal their electoral preferences plan to vote differently from those who do not express a preference. The use of multiple imputation (hypothesis 6) can be a way of correcting for these possible deviations (King, 2001; Liu, 2014). As a result, two different imputation techniques have been proposed, univariate and chained, and two sets of variables, one basic (sex, age, education, past vote and ideology) and the other enhanced (also including evaluation of the government and the economic situation).

GRAPH 3. Comparison of the Weighted Mean Absolute Error (WMAE) for estimating the imputed vote in its different versions in comparison with the base estimate (BE) without imputation



The reference for comparison is the base estimate (BE).

The use of univariate imputation (graph 3.a) or chained imputation (graph 3.c) employing a basic set of variables does not improve the estimates. Only in the case of expanded versions (graphs 3.b and .c) – which include survey participants' assessment of government action and the economic situation and which can only be computed starting with the 2000 elections –, do we find a slight improvement in the accuracy of estimates (simple: $F_{(1, 318)} = 12.97; p = 0.002$; chained: $F_{(1, 318)} = 8.51; p = 0.015$). Liu (2014), along the same lines, using a similar methodology to correct for voting intention in elections in Taiwan, shows the null capacity of this technique to correct for possible bias in voting intention. However, with the data available from the CIS, we can conclude that the low effectiveness of the imputation model is caused by the inadequate choice of predictors.

Thus, in the case of expanded imputation, we find a positive effect, which seems to be reflected in the 2004 and 2011 elections. However, it cannot be shown that there is a clear tendency related to whether the elections are ones in which there is a change in government (hypothesis 7), as occurs in the case of the use of weightings. In part this is because the magnitude of the improvement in the accuracy of the estimates using expanded imputation is minimal.

CONCLUSIONS

The first hypothesis raised in this study addresses the effectiveness of the use of socio-demographic weighting. The results show that despite the fact that the use of weighting produces a limited improvement in estimates until 1993, the contribution has been null since the mid-1990s. This is consistent with what has been found in other case studies (Crespi, 1988; Durand *et al.*, 2015). Either because socio-demographic variables are increasingly less and less related to voting or because the samples are rel-

atively balanced in this regard, the effect of this type of weighting on voting estimates is very limited. As suggested in Sturgis *et al.* (2016), one of the keys is to use population characteristics that are capable of correcting deviations resulting from non-response.

This study, in line with others previously cited, has shown the positive effect of weighting by past vote in the Spanish case. Thus, the second hypothesis is confirmed. However, even if the sample is balanced with respect to the preferences of voters in the most recent election, for this weighting to be successful it is necessary that the voters for a specific party that respond to the survey be representative of those not participating. This requirement is never fully met, and therefore the weighting by recall of a past vote is not sufficient to completely eliminate the bias present in the vote estimate.

In addition, the past vote variable itself may have deficiencies, due to respondents' failures of memory or to the social desirability of the response in certain contexts (Crespi, 1988). To correct these possible problems, two hypotheses were put forward in this paper: The first is that the combination of socio-demographic variables with past vote should be the most effective way of weighting the estimate; the second is that the correction of possible defects in the past vote variable could help to improve the quality of the adjustment. Neither of these hypotheses has been confirmed. What has been determined is that multiple imputation, in the way it has been used, has only lightly corrected the bias present in the past vote variable. This result, rather than raising doubts about imputation as a technique, reveals the importance of the predictors that are selected (Mercer *et al.*, 2018).

In addition, the use of imputation, as stated in the fourth hypothesis, has not resulted in a substantial improvement in the accuracy of the estimates. However, one significant contribution is that in the expanded imputation models, where predictors are added re-

garding respondents' assessment of the government and the economic situation, the ability to reduce the error in the estimate is greater. In line with the findings of Rivas *et al.* (2010), this points to the fact that the choice of predictors is the key, which also applies to the imputation of past vote.

The positive effect of weighting, and more tentatively, of enhanced imputation, is not consistent over time. This trend had already been detected by Escobar *et al.* (2014): in elections that bring about political change, the effect of weighting by past vote is nil, and in some cases, counter productive. This trend is also reflected in this study with the inclusion of data for 2015 and 2016. In the 2015 elections, which are considered elections of political change, the use of the weighting by past vote has a null effect, a situation that was partially reversed in 2016, elections in which the results of 2015 were largely reproduced. In the case of enhanced imputation, no conclusive trend is observed, although a positive effect is concentrated in the 2004 and 2011 studies, elections of political change in which weighting by past vote does not work correctly.

It would make sense in the future to abandon pre-election polls due to the limited number of questions they include, and it would be advisable to study which variables included in post-election studies could yield better results in improving estimates of the electoral behaviour of citizens through surveys.

BIBLIOGRAPHY

Abrams, Mark (1970). "The Opinion Polls and the 1970 British General Election". *Public Opinion Quarterly*, 34(3): 317-324.

Abramson, Paul R. (2007). "The French Presidential Election of 2007: Was Sarkozy the Condorcet Winner?". *French Politics* 5(3):287-291.

Anderson, Leslie (1992). "Surprises and Secrets: Lessons from the 1990 Nicaraguan Election". *Studies In Comparative International Development*, 27(3):93-119.

Ansolabehere, Stephen and Hersh, Eitan (2012). "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate". *Political Analysis*, 20(4):437-459.

Bernhagen, Patrick and Marsh, Michael (2007). "The Partisan Effects of Low Turnout: Analyzing Vote Abstention as a Missing Data Problem". *Electoral Studies*, 26(3):548-560.

Bethlehem, Jelke; Cobben, Fannie and Schouten, Barry (2011). *Handbook of Nonresponse in Household Surveys*. New York: John Wiley & Sons.

Biemer, Paul (2010). "Total Survey Error: Design, Implementation, and Evaluation". *Public Opinion Quarterly*, 74(5):817-848.

Biemer, Paul and Lyberg, Lars E. (2003). *Introduction to Survey Quality*. New York: John Wiley & Sons.

Bosch, Agustí, and Riba, Clara (2005). "Coyuntura económica y voto en España". *Revista de Sociología*, 75: 117-140.

Caballé, Adriá; Grima, Pere and Marco-Almagro, Lluís (2013). "Are Election Polls Right? Analysis of the Accuracy of Election Polls Predictions Published in the Press". *Revista Española de Investigaciones Sociológicas*, 143:25-46.

Callegaro, Mario and Gasperoni, Giancarlo (2008). "Accuracy of Pre-Election Polls for the 2006 Italian Parliamentary Election: Too Close to Call". *International Journal of Public Opinion Research* 20(2):148-170.

Crespi, Irving (1988). *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation.

Crewe, Ivor (1997). "The Polls: Confidence Restored?" *Parliamentary Affairs*, 50: 569-585.

Crewe, Ivor (2001). "The Opinion Polls: Still Biased to Labour". *Parliamentary Affairs*, 54(4):650-665.

Curtice, John (1997). *So How Well Did They Do? The Polls in the 1997 Election*. Centre for Research into Elections and Social Trends. London: Centre for Research into Elections and Social Trends.

D'Souza, John (2010). *Calibrate: a Stata Program for Calibration Weighting*. London: Stata User Group.

D'Souza, John (2011). *Calibrate: Stata module to calibrate survey datasets to population totals. Statistical Software Components S457240*. Boston College Department of Economics.

- Durand, Claire (2008). "The Polls of the 2007 French Presidential Campaign: Were Lessons Learned from the 2002 Catastrophe?". *International Journal of Public Opinion Research*, 20(3):275-298.
- Durand, Claire; Blais, André and Larochelle, Mylène (2004). "The Polls - Review. The Polls in the 2002 French Presidential Election: An Autopsy". *Public Opinion Quarterly* 68(4):602-622.
- Durand, Claire; Blais, André and Vachon, Sébastien (2001). "A Late Campaign Swing or a Failure of the Polls? The Case of the 1998 Quebec Election". *Public Opinion Quarterly*, 65(1):108-123.
- Durand, Claire; Deslauriers, Melanie and Vallois, Isabelle (2015). "Should Recall of Previous Votes Be Used to Adjust Estimates of Voting Intention?". *Survey Methods: Insights from the Field* 1-14. Available at: <http://surveyinsights.org/?p=3543>, access September 19, 2017
- Escobar, Modesto and Jaime, Antonio M. (2013). "Métodos de Imputación Múltiple para Predecir Resultados Electorales". In: Mendoza Velázquez, A. (ed.). *Aplicaciones en Economía y Ciencias Sociales con Stata*. Texas: Stata Press.
- Escobar, Modesto; Rivière Gómez, Jaime and Cilleros Conde, Roberto (2014). *Los Pronósticos Electorales con Encuestas: Elecciones Generales en España (1979-2011)*. Madrid: Centro de Investigaciones Sociológicas.
- Fuente, Angel de la and Domenech Vilarino, Rafael (2015). *El Nivel Educativo de la Población en España y sus Regiones: 1960-2011*. BBVA Bank, Economic Research Department. Madrid: BBVA Research. Available at: https://www.bbvaresearch.com/wp-content/uploads/2015/02/WP_15-07_Educacion.pdf, access September 19, 2017.
- Fuente, Angel de la (2015). *Series Enlazadas de los Principales Agregados Nacionales de la EPA, 1964-2014*. Instituto de Análisis Económico (CSIC). Madrid: FEDEA Research. Available at: <http://documentos.fedea.net/pubs/eee/eee2015-07.pdf>, access September 19, 2017.
- Groves, Robert M. et al. (2013). *Survey Methodology*. New York: Wiley.
- Hyndman, Rob J. and Koehler, Anne B. (2005). "Another Look at Measures of Forecast Accuracy," *Monash Econometrics and Business Statistics Working Papers* 13/05. Monash University, Department of Econometrics and Business Statistics.
- Jaime Castillo, Antonio M., and Sáez Lozano, José L. (2001). *El comportamiento electoral en la democracia española*. Madrid: Centro de Estudios Políticos y Constitucionales.
- Jennings, Will, and Wlezien, Christopher (2018). "Election polling errors across time and space". *Nature Human Behaviour* 1.
- Jowell, Roger et al. (1993). "Review: The 1992 British Election: The Failure of the Polls". *Public Opinion Quarterly*, 57(2):238-263.
- Katz, Daniel (1941). "The Public Opinion Polls and the 1940 Election". *Public Opinion Quarterly*, 5(1):52-78.
- King, Gary et al. (2001). "Analyzing Incomplete Political Science Data". *American Political Science Review*, 85(1269):49-69.
- Lago Peñas, Ignacio, and Lago Peñas, Santiago (2005). "Does the economy matter? An empirical analysis of the causal chain connecting the economy and the vote in Galicia". *Economics and Politics*, 17: 215-243.
- Lewis-Beck, M. S. (2005). "Election Forecasting: Principles and Practice". *The British Journal of Politics & International Relations*, 7: 145-164.
- Liu, Frank C. S. (2014). "Using Multiple Imputation for Vote Choice Data: A Comparison across Multiple Imputation Tools". *Open Journal of Political Science*, 4:39-46.
- Lundström, Sixten and Särndal, Carl E. (2001). *Estimation in the Presence of Nonresponse and Frame Imperfection*. Suecia: Statistics Sweden.
- Lynn, Peter (2008). "The Problem of Nonresponse". In: European Association of Methodology (ed.). *The International Handbook of Survey Methodology*. New York: Lawrence Erlbaum Associates.
- Lynn, Peter and Jowell, Roger (1996). "How Might Opinion Polls Be Improved? The Case for Probability Sampling". *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159(1):21-28.
- McEwen, Nicola (2004). "Opinion Polling in Scotland: An Analysis of the 2003 Scottish Parliament Election". *Journal of Elections, Public Opinion and Parties*, 14(1):171-190.
- Mercer, Andrew; Lau, Arnold and Courtney, Kennedy (2018). *For Weighting Online Opt-In Samples, What Matters Most?*. Washington: Pew Research Centre.
- Orriols, Lluís and Cordero, Guillermo (2016). "The Breakdown of the Spanish Two-Party System:

- The Upsurge of Podemos and Ciudadanos in the 2015 General Election". *South European Society and Politics*, 21:4, 469-492.
- Pavía, Jose M.; Badal, Elena and García-Cárceles, Belén (2016). "Spanish Exit Polls. Sampling Error or Nonresponse Bias?". *Revista Internacional de Sociología*, 74(3):e043
- Pavía, José M. and Larraz, Beatriz (2012). "Sesgo de no-respuesta y modelos de superpoblación en encuestas electorales". *Revista Española de Investigaciones Sociológicas*, 137:121-150.
- Rama, José (2016). *Crisis económica y sistema de partidos: síntomas de cambio político en España*. Barcelona: Institut de Ciències Polítiques i Socials.
- Rivas, Cristina; Martínez Rosón, María del Mar and Galindo, Purificación (2010). "La Imputación Múltiple como Alternativa al Análisis de la No Respuesta en la Variable Intención de Voto". *Revista Española de Ciencia Política*, 22, 99-118.
- Rivero Rodríguez, Gonzalo (2011). *Análisis de datos incompletos en Ciencias Sociales*. Madrid: CIS.
- Rosenthal, Robert; Rosnow, Ralph L. and Donald B. Rubin (2000). *Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach*. Cambridge: Cambridge University Press.
- Rubin, Donald B. (1987). *Multiple Imputation for Non-response in Surveys*. New York: John Wiley & Sons.
- Sanders, David (2003). "Pre-Election Polling in Britain, 1950-1997." *Electoral Studies*, 22(1):1-20.
- Särndal, Carl E. (2007). "The Calibration Approach in Survey Theory and Practice". *Survey Methodology*, 33(2):99-119.
- Särndal, Carl E. and Lundström, Sixten (2005). *Estimation in Surveys with Nonresponse*. England: John Wiley & Sons.
- Sauger, Nicolas (2008). "Assessing the Accuracy of Polls for the French Presidential Election: The 2007 Experience". *French Politics*, 6(2):116-136.
- Sciarini, Pascal and Goldberg, Andreas C. (2016). "Turnout Bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting". *Journal of Survey Statistics and Methodology*, 4(1):110-137.
- Shlapentokh, Vladimir (1994). "The Polls - a Review - the 1993 Russian Election Polls". *Public Opinion Quarterly*, 58(46302):579-602.
- Smith, Fred T. M. (1996). "Public Opinion Polls: The UK General Election, 1992". *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(3):535-545.
- Stata (2015). *Stata 14 Base Reference Manual*. College Station, Texas: Stata Press.
- StataCorp (2017). *Stata Multiple Imputation Reference Manual*. Texas: StataCorp.
- Sturgis, Patrick; Allum, Nick and Brunton-Smith, Ian (2009). "Attitudes Over Time: The Psychology of Panel Conditioning". In: Groves, R. M.; Kalton, G.; Rao, J. N.; Schwarz, N.; Skinner C. and Lynn, P. (eds.). *Methodology of Longitudinal Surveys*. Wiley: Nueva York.
- Sturgis, Patrick et al. (2016). *Report of the Inquiry into the 2015 British General Election Opinion Polls*. British Polling Council: London.
- Traugott, Michael W. (2005). "The Accuracy of the National Preelection Polls in the 2004 Presidential Election". *Public Opinion Quarterly*, 69(5 SPEC. ISS.):642-654.
- Urquiza Sancho, Ignacio (2005). "El Voto Oculto en España". *Revista Española de Ciencia Política*, 13:119-156.
- Varela Mallou, Jesús et al. (1998). "Estimación de la Respuesta de los 'No Sabe/No Contesta' en los Estudios de Intención de Voto". *Revista Española de Investigaciones Sociológicas*, 83:269-287.
- Voogt, Robert J. J. and Saris, William E. (2003). "To Participate or Not to Participate: The Link Between Survey Participation, Electoral Participation, and Political Interest". *Political Analysis* 11(2):164-179.
- Voss, Stephen; Gelman, Andrew and King, Gary (1995). "Preelection Survey Methodology: Details from Eight Polling Organizations, 1988 and 1992". *Public Opinion Quarterly*, 59:98-132.
- Worcester, Robert (1996). "Political Polling: 95% Expertise and 5% Luck." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(1):5.

RECEPTION: September 25, 2017

REVIEW: February 14, 2018

ACCEPTANCE: May 18, 2018

Appendix 1. Regression models

TABLE 1A. Linear regression models

	M1	M2	M3	M4	M5
Elections					
1982					
1986	1.16 (0.81)				-1.41 (0.96)
1989	1.04 (0.81)				0.93 (0.96)
1993	-1.40 (0.81)				1.15 (0.96)
1996	-3.44*** (0.81)				-3.53*** (0.96)
2000	-2.84*** (0.81)				-5.38*** (0.96)
2004	-1.93 (1.00)	1.04* (0.48)			-1.04 (1.18)
2008	0.37 (0.81)	3.18*** (0.39)			-2.14* (0.96)
2011	-4.85*** (0.81)	-1.78*** (0.39)			-5.36*** (0.96)
2015	-2.56** (0.81)	0.53 (0.39)			-2.55* (0.96)
2016	-2.29** (0.81)	0.73 (0.39)			-4.06*** (0.96)
Political change					
Elections of continuity					
Elections of change			-1.97*** (0.47)	-1.62*** (0.36)	
Weighting					
Base estimate (BE)					
SD	-0.87 (0.72)	0.56 (0.31)	-0.75 (0.39)	-0.27 (0.27)	
PV0	-1.06 (0.72)	-0.87** (0.31)	-3.26*** (0.39)	-2.09*** (0.27)	
PV1	0.34 (0.72)	-0.51 (0.31)	-3.10*** (0.39)	-1.74*** (0.27)	
PV2	1.65* (0.72)	-1.14*** (0.31)	-3.05*** (0.39)	-1.93*** (0.27)	
PV3	1.49* (0.72)	-1.33*** (0.31)	-3.18*** (0.39)	-2.00*** (0.27)	
SD + PV	0.77 (0.72)	-1.31*** (0.31)	-3.15*** (0.39)	-1.94*** (0.27)	

TABLE 1A. *Linear regression models* (continuation)

	M1	M2	M3	M4	M5
Imputation methods					
Base estimate (BE)					
Basic univariate	-0.82 (0.47)	0.11 (0.26)	0.03 (0.25)	-0.05 (0.23)	
Enhanced univariate		-0.34 (0.26)		-0.18 (0.23)	
Basic chained	-0.63 (0.47)	-0.17 (0.26)		-0.13 (0.23)	
Enhanced chained		-0.30 (0.26)		-0.17 (0.23)	
Type of study					
Pre-electoral	-0.53***	-0.67	-0.62***	-0.84***	0.44
Post-electoral	(0.12)	(0.07)	(0.16)	(0.11)	(0.25)
Treatment of PV					
PV1 without imputation					
PV2 imputed (NR)					2.07* (0.96)
PV3 imputed (NR y NVJ)					1.64 (0.96)
Constant	6.23*** (0.58)	3.26*** (0.28)	5.61*** (0.32)	4.65*** (0.25)	10.34*** (0.69)
F	7.66	16.66	11.71	11.35	11.24
Degrees of freedom	99	66	18	22	33
P-value	0.000	0.000	0.000	0.000	0.000
R squared	0.60	0.73	0.30	0.37	0.85
Casos	441	385	441	385	63

M1: Weighted Mean Absolute Error (WMAE) for estimated vote in the period 1982-2016 only includes imputation methods with basic variables.

M2: Weighted Mean Absolute Error (WMAE) for estimated vote in the period 2000-2016 includes all of the imputation methods.

M3: M1 with the variable Elections substituted by an indicator of political change (Elections of change: 1982, 1996, 2004, 2011 and 2015).

M4: M2 with the variable Elections substituted by an indicator of political change.

M5: WMAE of the past vote variable for the period 1982-2016.

Weightings. SD: Sociodemographic variables; PV0: past vote filtering NR/NA and did not vote;

PV1: past vote without filtering;

PV2; by past vote with imputed NR and NA;

PV3 by past vote NR, NA and without imputed age; PV+SD: for sociodemographic variables and PV3.

Treatment of PV (past vote): PV1 without imputation; PV2 imputed (NR): past vote imputed to those who did not respond;

PV3 imputed (NR and NVM): past vote imputed to those that did not respond and that were not of voting age.

* p<0.05, ** p<0.01, *** p<0.01.

El efecto de la ponderación y la imputación en el sesgo de los estudios electorales en España

The Effect of Weighting and Multiple Imputation on Bias in Spanish Election Polls

Pablo Cabrera-Álvarez y Modesto Escobar

Palabras clave

- Encuestas
- Estimación de voto
- Imputación múltiple
- Recuerdo de voto
- Ponderación
- *Total Survey Error*

Key words

- Polls
- Voting Estimates
- Multiple Imputation
- Past Vote
- Weighting
- Total Survey Error

Resumen

Este artículo tiene como objetivo evaluar la eficacia de las correcciones realizadas en encuestas electorales para ajustar el efecto de la no respuesta. Para ello se ponen a prueba distintos métodos de ponderación e imputación múltiple en todos los estudios preelectorales y postelectorales de elecciones generales al Congreso de los Diputados español llevados a cabo por el Centro de Investigaciones Sociológicas desde 1982. Los resultados muestran la ventaja de utilizar la variable recuerdo de voto en la ponderación cuando hay estabilidad en las preferencias de los electores. De modo complementario, para tratar la no respuesta, el uso de técnicas de imputación tiene un efecto limitado y condicionado por las variables incluidas en el modelo.

Abstract

The purpose of this article is to assess the effectiveness of post-survey adjustments made to electoral polls in order to correct for non-response bias. To do so we have used different weighting and multiple imputation methods using pre-election and post-election polls conducted by Spain's Centre for Sociological Research for all Spanish general elections since 1982. The results show the benefit of weighting by past vote when voters' preferences remain stable. However, the use of multiple imputation techniques to address missing values has a limited effect and is influenced by the variables included in the model.

Cómo citar

Cabrera-Álvarez, Pablo y Escobar, Modesto (2019). «El efecto de la ponderación y la imputación en el sesgo de los estudios electorales en España». *Revista Española de Investigaciones Sociológicas*, 165: 45-64. (<http://dx.doi.org/10.5477/cis/reis.165.45>)

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

Pablo Cabrera-Álvarez: Universidad de Salamanca | pablocal@usal.es

Modesto Escobar: Universidad de Salamanca | modesto@usal.es

INTRODUCCIÓN¹

Las estimaciones de voto hechas a partir de encuestas preelectorales tienden a aproximarse al resultado final de las elecciones según se va acercando la cita con las urnas. Aunque en los últimos años hay estudios que apuntan a que la precisión de las encuestas preelectorales se ha mantenido en los límites de lo esperado (Jennings y Wlezien, 2018), también hemos asistido a casos, como los de las elecciones generales del Reino Unido de 2015 y las elecciones generales de España de 2016, que han causado un amplio debate en la esfera pública sobre la bondad y necesidad de las encuestas preelectorales.

Entre las causas que explican la falta de precisión de las encuestas preelectorales se encuentra la existencia del sesgo de no respuesta, es decir, la diferencia sistemática entre la intención de voto de aquellos que responden a la encuesta y la de quienes no responden. Para mitigar este sesgo existen técnicas estadísticas que corrigen posibles desviaciones en el perfil de la muestra empleando información complementaria. Una de las variables auxiliares habitualmente utilizadas en este procedimiento es el recuerdo de voto, es decir, el comportamiento en las anteriores elecciones declarado por los entrevistados. A pesar de ello, sigue existiendo debate sobre el empleo de esta variable, ya que no hay ninguna evidencia de que siempre tenga un efecto positivo sobre la precisión de las estimaciones de voto. Con el mismo objetivo de reducir el sesgo producido por la no respuesta, se pueden utilizar técnicas de imputación múltiple (IM) para asignar valores válidos a los que dicen no saber o no responden a la pregunta sobre la intención de voto.

Este artículo, centrado en el caso español, busca determinar el efecto del uso de la

imputación múltiple y la ponderación sobre la precisión de la estimación de voto, tomando una perspectiva temporal que compara encuestas pre y postelectorales de las elecciones al Congreso de los Diputados en el período 1982-2016.

Frente a otros trabajos previos (Escobar *et al.*, 2014; Pavía y Larraz, 2012; Rivas *et al.*, 2010), la relevancia y novedad de este radica en tres razones. La primera es el uso de diferentes transformaciones de la variable recuerdo de voto en la ponderación, así como de un conjunto de variables sociodemográficas que no habían sido probadas hasta ahora en España. La segunda es la utilización de la imputación múltiple de modo extenso para tratar tanto la intención de voto como el recuerdo de voto. La tercera es que tanto las ponderaciones como las imputaciones son testadas en estudios pre y postelectorales del Centro de Investigaciones Sociológicas (CIS), cubriendo las elecciones generales celebradas en España desde 1982. Esta perspectiva temporal es necesaria para aclarar si el funcionamiento de estas técnicas es cambiante de una elección a otra, y si existe alguna tendencia que apunte al porqué de esa variabilidad.

El texto que sigue a continuación se divide en cuatro apartados. En el primero se aborda una definición de precisión y se plantea el paradigma del *Total Survey Error* (TSE) como la teoría marco para abordar el estudio de encuestas electorales. A continuación, se presentan las hipótesis de trabajo, y posteriormente se exponen en detalle los datos y la metodología empleados. A este apartado le siguen los resultados del análisis y la discusión. El último apartado está dedicado a la conclusión.

PRECISIÓN Y FUENTES DE ERROR EN LAS ENCUESTAS PREELECTORALES

El concepto de precisión en las encuestas electorales hace referencia a dos cuestiones.

¹ Esta investigación ha sido apoyada por el programa de becas predoctorales de la Obra Social «La Caixa».

Por un lado, a la noción de variabilidad derivada del tamaño muestral y la varianza de la estimación, representado por el margen de error muestral, y por el otro, a la noción de ajuste, como diferencia entre la estimación y los datos poblacionales, en este caso, el resultado de las elecciones. En los trabajos, como este, que analizan el desempeño de las encuestas, se utiliza el segundo concepto teniendo en cuenta que esa comparación puede verse alterada por los efectos de la campaña (Crespi, 1988; Sturgis *et al.*, 2016).

La falta de precisión de las encuestas preelectorales es un tema recurrente en la literatura sobre opinión pública y elecciones (Caballé *et al.*, 2013; Callegaro y Gasperoni, 2008; Durand *et al.*, 2004; Sanders, 2003; Traugott, 2005). Entre las causas de estos desajustes se encuentran el cambio de preferencias de los votantes entre la realización de la encuesta y las elecciones (Abrams, 1970; Shlapentokh, 1994), los métodos utilizados para determinar la probabilidad de ir a votar (Durand *et al.*, 2004; McEwen, 2004; Sturgis *et al.*, 2016), el muestreo empleado (Abramson, 2007; Curtice, 1997; Lynn y Jowell, 1996; Worcester, 1996), los problemas de cobertura de la población (Callegaro y Gasperoni, 2008; Durand *et al.*, 2001; Sauger, 2008) o el tratamiento de los casos que no responden a la encuesta o a la pregunta de intención de voto (Anderson, 1992; Jowell *et al.*, 1993; Katz, 1941).

El marco del TSE permite analizar de forma sistemática las fuentes de error que existen en el proceso de diseño, recogida, procesamiento y análisis de los datos de una encuesta. Conocer y controlar estas fuentes de error es fundamental para mantener la calidad de las estimaciones (Biemer, 2010; Biemer y Lyberg, 2003). El TSE divide las fuentes de error en dos grupos, aquellas relacionadas con la medición, y las relacionadas con la representatividad. En el lado de la medición se encuentran la validez, el error de medición y el error de procesamiento de los datos. Por su parte, en el lado de la representatividad están

el error de cobertura, el error muestral, el error de no respuesta y el error derivado de los ajustes realizados tras la recogida de datos (Groves *et al.*, 2013).

El sesgo de no respuesta en encuestas preelectorales

Este trabajo se centra en el error de no respuesta. La no respuesta se refiere a la falta de información debida a que el elemento muestral no sea localizado o rechace colaborar con la encuesta o una parte de esta (Lynn, 2008). En el contexto de este estudio decimos que existe un sesgo de no respuesta cuando los que responden a la encuesta, o a la pregunta sobre intención de voto, tienen preferencias electorales diferentes de los que no responden.

La no respuesta total, es decir, cuando el elemento muestral rechaza participar en la encuesta o no es contactado, ha sido señalada como una de las causas de los problemas de precisión de las encuestas preelectorales en diferentes países, entre ellos España (Durand *et al.*, 2004; Jowell *et al.*, 1993; Smith, 1996; Pavía *et al.*, 2016). En esta dirección, algunos estudios han demostrado que la propensión a responder a encuestas sociopolíticas está relacionada con el nivel de interés por la política de los ciudadanos seleccionados en la muestra (Voogt y Saris, 2003). Más recientemente se ha comprobado que aquellos que van a votar tienden a estar sobrerrepresentados en estudios postelectorales, contribuyendo a que la estimación de la participación realizada a partir de la encuesta exceda la cifra real (Ansolabehere y Hersh, 2012; Sciarini y Goldberg, 2016).

El otro fenómeno asociado con el sesgo de no respuesta es la negativa de algunos participantes a desvelar su intención de voto. Este sesgo también ha sido apuntado como una posible causa de la falta de precisión de las encuestas preelectorales (Curtice, 1997; McEwen, 2004; Sauger, 2008). Referido a

España, Urquizu (2005) demostró que los votantes conservadores eran menos dados a desvelar sus preferencias electorales durante la década de los ochenta, mientras que en los noventa esta tendencia se invirtió.

Ponderación e imputación como métodos para ajustar el sesgo de no respuesta en estudios electorales

Una vez que los datos han sido recogidos es posible aplicar ajustes para reducir el impacto de los sesgos causados por la no respuesta total y parcial. En el caso de la no respuesta total se emplean técnicas de ponderación para reequilibrar la muestra final. De la misma forma, también se pueden utilizar técnicas de imputación para atribuir categorías de respuesta válidas a aquellos que evitan contestar alguna pregunta.

El uso de ponderaciones por calibración y técnicas similares ha sido habitual en combinación con el muestreo por cuotas (Särndal, 2007). A pesar de ello, en la literatura existe un debate sobre la conveniencia de incluir el recuerdo de voto como una variable auxiliar en la ponderación. En Estados Unidos, por ejemplo, la mayoría de las empresas demoscópicas han evitado el uso de esta variable (Voss *et al.*, 1995). En una clásica investigación sobre la metodología de las encuestas preelectorales en Estados Unidos, Crespi (1988: 40-41) establece que el principal motivo esgrimido para no utilizar esta variable son los problemas de medición que tiene asociados, entre los que destaca la sobrerrepresentación de los votantes frente a los abstencionistas y la de los que optan por el partido o candidato ganador frente a los que votan a los perdedores. Worcester (1996), refiriéndose al caso del Reino Unido, alegó que su utilización, lejos de ayudar, podía conducir a estimaciones menos precisas. En la misma línea, en una investigación reciente sobre encuestas preelectorales en Canadá y Francia, Durand *et al.* (2015) han demostrado que el uso del voto pasado pue-

de tanto mejorar como empeorar la precisión de las estimaciones.

A pesar de este debate, el uso de la variable recuerdo de voto en la ponderación es habitual. Por ejemplo, en el Reino Unido y en Francia esta variable se ha utilizado para corregir la subrepresentación de los conservadores (Crewe, 2001) y del Frente Nacional (Durand, 2008), respectivamente. En España, solo algunos estudios académicos han intentado arrojar luz sobre este fenómeno. Escobar *et al.* (2014) compararon diferentes metodologías para realizar estimaciones de voto a partir de encuestas utilizando como variable de ponderación el recuerdo de voto. En ese trabajo encontraron que, en las elecciones en las que se produce cambio de partido ganador, el uso de la variable recuerdo de voto empeora la precisión de la estimación en el período 1979-2011. Por su parte, Pavía y Larraz (2012) también experimentaron con diferentes formas de ponderación por recuerdo de voto, llegando a la conclusión de que la posestratificación no era el método de ajuste más recomendable en presencia de sesgo de no respuesta.

En cuanto a la no respuesta parcial, una vez que los datos han sido recogidos, los investigadores tienen que afrontar la decisión de cómo proceder con los casos que, siendo probable que acudan a las urnas, no desvelan su intención de voto. Este problema ha sido abordado utilizando técnicas *ad hoc* definidas por cada organización (Crespi, 1988; Lynn y Jowell, 1996; Sturgis *et al.*, 2016). En España podemos encontrar el trabajo de Varela *et al.* (1998), que describe diferentes métodos que podrían ser utilizados para imputar una respuesta válida a los que no revelan sus preferencias electorales, así como el de Pavía y Larraz (2012), en el que se emplea imputación por criterio experto para tratar la no respuesta en la intención y el recuerdo de voto. También en el trabajo de Rivas *et al.* (2010) se discute la pertinencia del uso de la imputación para tratar la no respuesta parcial en la intención

de voto tomando el caso de las elecciones de 2000, la conclusión es que esta técnica es solo efectiva cuando los predictores permiten diferenciar a todas las categorías de la intención de voto.

La imputación múltiple es una técnica que se utiliza para asignar valores válidos a los casos que presentan valores perdidos, pero su aplicación a los estudios preelectorales ha sido reducida. King *et al.* (2001) analizaron su uso potencial en el ámbito de la ciencia política, señalando que podría utilizarse para el estudio de las preferencias de los no votantes. Bernhagen y Marsh (2007) utilizaron esta técnica para asignar un valor válido a aquellos que no declararon su intención de voto y Liu (2014) utilizó técnicas de imputación múltiple para asignar preferencias a aquellos que rechazaron desvelar su intención de voto en un estudio preelectoral sobre las elecciones en Taiwán sin éxito. Referido a España, Escobar y Jaime (2013) tampoco encontraron una mejora de la precisión de las estimaciones al usar diferentes métodos de imputación múltiple en los estudios pre y postelectoral del Centro de Investigaciones Sociológicas de las elecciones generales de 2011.

HIPÓTESIS DE INVESTIGACIÓN

Para conseguir el objetivo de determinar el efecto del uso de ponderación y de la imputación múltiple en la precisión de la estimación de voto en el caso español se han formulado las siguientes hipótesis:

Hipótesis 1: La ponderación de la muestra utilizando variables sociodemográficas afecta de forma positiva a la precisión de la estimación de voto.

Sin embargo, cabe esperar que este efecto sea limitado, porque en el muestreo ya suelen emplearse cuotas y porque generalmente este tipo de variables están poco relacionadas con la intención de voto y las técnicas de ponderación reducen el sesgo

de no respuesta total en la estimación cuando la información empleada está correlacionada con la propensión a responder la variable de interés, en este caso, la intención de voto (Särndal, 2007).

Hipótesis 2: La ponderación de la muestra por recuerdo de voto contribuye a incrementar la precisión de la estimación de voto.

Distinto es el caso de la variable recuerdo de voto, cuyo uso está bastante extendido en la industria de las encuestas de opinión, puesto que está relacionada con la intención de voto (Crespi, 1988; Crewe, 1997). Algunos trabajos académicos, sin embargo, han demostrado que el efecto de utilizar esta variable en otros países es mínimo (Durand, Deslauriers y Valois, 2015).

Hipótesis 3: El uso de técnicas de imputación múltiple para tratar la variable recuerdo de voto reduce el nivel de sesgo presente en esta variable y, en consecuencia, su uso en la ponderación de la variable recuerdo de voto imputada incrementa la precisión de la estimación de voto.

En la medida en que la variable de recuerdo de voto en las elecciones anteriores podría estar afectada por problemas de memoria o por un sesgo de no respuesta parcial (Crespi, 1988; Worcester, 1996), aquí se propone un procedimiento para corregir la desviación derivada de la no respuesta parcial mediante el uso de técnicas de imputación múltiple.

Hipótesis 4: El uso de la ponderación que combina las variables sociodemográficas y el recuerdo de voto será el más efectivo para reducir el nivel de error de la estimación de voto ($RV+SD$).

Esta hipótesis es un mero corolario de las precedentes. Si la ponderación con variables sociodemográficas y el recuerdo de voto afecta positivamente de modo separado, es de esperar que el resultado mejore si se combinan.

Hipótesis 5: *El uso de la variable recuerdo de voto para ponderar tiene un efecto positivo en la precisión de las estimaciones en las elecciones en las que hay continuidad política.*

En Escobar *et al.* (2014) se señala una tendencia sobre el uso del recuerdo de voto en las ponderaciones en el caso español (1979-2011): cuando en unas elecciones el gobierno vigente no consigue revalidar su hegemonía, el uso de la variable recuerdo de voto en la ponderación tiene un efecto negativo en la precisión. Como, tras 2014, el sistema de partidos en España cambió con la irrupción de dos nuevos partidos políticos y la disminución del porcentaje de apoyo de las dos primeras fuerzas políticas (Orriols y Cordero, 2016; Rama, 2016), cabe comprobar si esa generalización se ha cumplido en las dos siguientes elecciones (2015 y 2016).

Hipótesis 6: *El uso de técnicas de imputación múltiple para asignar una intención o conducta de voto a aquellos que no saben o no contestan incrementa la precisión de la estimación de voto.*

En este sentido, abogaríamos por el uso de la imputación múltiple para asignar valores válidos a aquellos que no desvelan su comportamiento de voto en las elecciones que se celebran en torno a la fecha de la encuesta (King, 2001).

Hipótesis 7: *Del mismo modo que ocurría con la ponderación con el recuerdo de voto, el uso de técnicas de imputación múltiple tiene efectos variables según el resultado de las elecciones sea progubernamental o no lo sea.*

Escobar y Jaime (2013) mostraron el efecto positivo de la imputación en las elecciones de 2011. Sin embargo, hasta la fecha no se ha publicado ningún otro estudio de los efectos de este procedimiento en otros comicios anteriores o posteriores.

METODOLOGÍA

En este apartado se presenta la metodología en cuatro secciones: en la primera se describen las fuentes de datos. A continuación, se abordan, por un lado, los diferentes criterios de ponderación empleados y, por el otro, en la tercera, los procedimientos de imputación utilizados. En la última parte se presentará el criterio empleado para determinar la precisión de las estimaciones de voto.

Muestra

Con el objetivo de ver las implicaciones de emplear diferentes tipos de ponderaciones e imputaciones se trabaja con las encuestas preelectorales y postelectorales elaboradas por el CIS entre 1982² y 2016 para las elecciones generales en España. Todas las muestras del CIS emplearon un muestreo polietápico estratificado por provincia y hábitat con selección de los hogares mediante rutas aleatorias y de los sujetos por cuotas de sexo y edad. La relación de estudios utilizados se muestra en la tabla 1.

La inclusión de los estudios postelectorales responde a la limitación que suponen las fechas del trabajo de campo de las encuestas preelectorales del CIS, en torno a un mes antes de la cita con las urnas. Con los estudios postelectorales se busca atenuar el posible sesgo introducido por los efectos de la campaña electoral, que no son detectados por las encuestas preelectorales. Sin embargo, también es necesario señalar otras limitaciones que presentan los estudios postelectorales: 1) los votantes en las elecciones están sobrerrepresentados (Sciarini y Goldberg, 2016); 2) en el caso de que los estudios sean de tipo panel, puede generarse un cier-

² No se incluyen las elecciones de 1977 y 1979 porque en ninguna de ellas está disponible un estudio postelectoral. Además, en el caso de las elecciones de 1977 no se dispone de recuerdo de voto, variable crucial en este estudio.

TABLA 1. Año, número y tamaño de los estudios del CIS empleados

Año	Números del CIS	Muestra preelectoral	Muestra postelectoral
1982	1.326 y 1.327	24.832	2.394
1986	1.526 y 1.542	25.304	6.842
1989	1.821/37 y 1.842	27.122	2.508
1993	2.060 y 2.061	2.462	4.225
1996	2.207 y 2.210	6.544	4.610
2000	2.382 y 2.384	24.040	4.386
2004*	2.555	24.109	
2008	2.750 y 2.757	18.221	5.247
2011	2.915 y 2.920	17.201	6.056
2015	3.117 y 3.126	17.403	5.457
2016	3.141 y 3.146	17.458	5.136

* El estudio postelectoral de 2004 ha sido excluido del análisis al no contar con la variable recuerdo de voto.

Nota: A partir de 2000, con la excepción de 2004 y 2016, los estudios pre y postelectorales fueron de tipo panel.

to condicionamiento sobre los entrevistados (Sturgis *et al.*, 2009), y 3) suele existir una sobrerrepresentación del partido ganador en las recientes elecciones (Crespi, 1988).

Crterios de ponderación

En los estudios preelectorales, el CIS diseña una muestra estratificada por provincia con afijación no proporcional. En estos casos, para realizar las estimaciones, se ha de aplicar un peso de selección. Estos pesos (w_k) son iguales para los entrevistados de una misma circunscripción y su fórmula viene dada por

$$w_k = e_k / n_k$$

siendo e_k el tamaño del censo electoral y n_k el número de entrevistas realizadas en cada estrato, provincia o circunscripción.

Para calcular el resto de ponderaciones se ha empleado el método de calibración logística. Existen otros métodos para generar coeficientes de ponderación como, por ejemplo, el uso de modelos de no respuesta para de-

terminar la probabilidad de responder a la encuesta o de métodos basados en técnicas de *propensity score matching*. La comparación de estos métodos muestra que la clave está en los predictores que se utilizan (Mercer *et al.*, 2018), más que en la técnica estadística empleada para generar las ponderaciones. En este caso, dado que los datos poblacionales solo se pueden obtener de forma agregada, la técnica empleada ha sido la calibración en su versión logística, cuya ventaja sobre la versión lineal es que evita la generación de coeficientes de ponderación negativos.

La calibración logística obliga, tras obtener la muestra, a cotejar la distribución de una o varias de sus variables para ver si coinciden con los parámetros de la población, a fin de calcular unos pesos que logren que los resultados muestrales coincidan con los poblacionales en las variables seleccionadas³. El primer criterio de calibra-

³ Para más información acerca del cálculo de coeficientes de ponderación por calibración se pueden consultar

ción empleado incluye únicamente variables sociodemográficas, mientras que el resto de ponderaciones contempla el recuerdo de voto⁴. En la tabla 2 se presenta un resumen de las ponderaciones empleadas.

Criterios de imputación

Se ha empleado el método propuesto por Rubin (1987) para el análisis e imputación de datos incompletos, que consiste en reconstruir nuevos conjuntos de datos, tantos como el investigador establezca, con valores simulados aleatoriamente por otras variables del estudio que contengan mayor información⁵. Frente a la imputación única, que consiste en estimar una sola vez los datos, la imputación múltiple realiza una serie de estimaciones —mediante simulación de un número de conjuntos de datos completos— a partir de las cuales se puede reconstruir una única estimación, complementada con la variación de las diversas estimaciones realizadas. En consecuencia, las varianzas de los parámetros pueden obtenerse de modo más certero que con la imputación única.

los trabajos de Särndal (2005) o Lundström y Särndal (2001). La calibración fue realizada en Stata utilizando el paquete *calibrate* diseñado por D'Souza (2011).

⁴ Para realizar la calibración es necesario contar con las distribuciones poblacionales de las variables auxiliares. Los datos para realizar la ponderación por recuerdo de voto y tamaño de hábitat proceden del Ministerio del Interior. La información sobre las distribuciones poblacionales de las variables sexo, edad y comunidad autónoma ha sido recogida del Instituto Nacional de Estadística (INE). Por su parte, los datos históricos sobre nivel educativo y estatus laboral fueron obtenidos de los trabajos de Fuente y Domenech (2015) y Fuente (2015), respectivamente. En el caso de las dos últimas elecciones, 2015 y 2016, para las que no existían datos en los trabajos anteriores, los valores poblacionales se obtuvieron de la encuesta de población activa (INE) en el caso del estatus laboral y fueron interpolados para el nivel educativo.

⁵ Una introducción básica junto con el modo de obtener estos modelos con Stata puede encontrarse en el volumen dedicado a la imputación múltiple (Stata, 2015). Asimismo, una presentación teórica y aplicada en español se encuentra en el ya citado libro de Rivero (2011).

Existen diferentes procedimientos de imputación basados en principios bayesianos y frecuentistas. Básicamente, pueden distinguirse, por un lado, las imputaciones univariadas (una sola variable al mismo tiempo) a partir de la distribución predictiva posterior de los datos perdidos, y, por el otro, las imputaciones encadenadas, que implican una imputación retroalimentada de las variables que imputan. La tabla 3 resume los diferentes tipos de imputaciones tenidas en cuenta para este trabajo (nótese la correspondencia entre las versiones del recuerdo de voto imputado y las ponderaciones con el mismo nombre).

La elección de los predictores incluidos en los modelos de imputación se hicieron atendiendo a criterios teóricos (Escobar y Jaime, 2013) y contando con las limitaciones derivadas del diseño de investigación. En primer lugar, se recurrió a estudios sobre comportamiento electoral con el fin de determinar qué predictores sociodemográficos y políticos están relacionados con la intención de voto (p. ej., Bosch y Riba, 2005; Jaime y Saéz, 2001; Lago y Lago, 2005). En segundo lugar, la lista fue limitada con el fin de poder aplicar el mismo modelo en toda la serie de elecciones estudiadas.

Dado que la variable de interés para la imputación es siempre el voto (x_i), se empleó un modelo multinomial en el que las probabilidades de las k categorías de la variable respondían a la siguiente fórmula, donde \mathbf{z}_i es el vector de variables empleadas para imputar:

$$\Pr(x_i = k | \mathbf{z}_i) = \frac{1}{1 + \sum_{l=2}^K \exp(\mathbf{z}_i' \boldsymbol{\beta}_l)}, \text{ si } k = 1$$

$$\Pr(x_i = k | \mathbf{z}_i) = \frac{\exp(\mathbf{z}_i' \boldsymbol{\beta}_k)}{1 + \sum_{l=2}^K \exp(\mathbf{z}_i' \boldsymbol{\beta}_l)}, \text{ si } k > 1$$

TABLA 2. Esquema de las ponderaciones por provincia, recuerdo de voto (RV) y variables sociodemográficas empleadas en el diseño de la investigación

Sigla	Variables de ponderación	RV imputado	Ámbito del RV*
EB	Provincia		
SD	Provincia y sociodemográficas**		
RV0	Provincia y recuerdo de voto (RV)	No	NR, NV y NVJ excluidos
RV1	Provincia y recuerdo de voto	No	NR, NV y NVJ incluidos
RV2	Provincia y recuerdo de voto	Sí (NR imputado)	NR, NV y NVJ incluidos
RV3	Provincia y recuerdo de voto	Sí (NR y NVJ imputado)	NR, NV y NVJ incluidos
SD+RV	Prov., sociodemográficas** y RV	Sí (NR y NVJ imputado)	NR, NV y NVJ incluidos

NR: no respuesta (NS y NC); NV: no votó en las anteriores elecciones; NVJ: no votó por ser menor de edad.

* A las categorías no imputadas, sean por no respuesta (NR) o por no votar debido a edad (NVJ), se les asigna solo el peso correspondiente al resto de criterios de ponderación.

** Edad por sexo, comunidad autónoma, tamaño de hábitat, nivel educativo y situación laboral.

Evaluación de la precisión de los resultados

Para evaluar el resultado de estimaciones e imputaciones se hace uso del error medio absoluto ponderado (EMAP), empleado en la literatura de pronóstico en las series temporales⁶, cuya fórmula viene dada por la siguiente expresión:

$$\epsilon = \sum_{k=1}^K |\hat{p}_k - p_k| p_k$$

donde p_k es el resultado electoral de cada k partido y \hat{p}_k sus correspondientes estimaciones.

⁶ Tres son las medidas más empleadas para estos propósitos: la media del error absoluto, la media del error cuadrático y la media del error cuadrático estandarizada. Por su visibilidad se ha empleado la primera, añadiendo la ponderación de los errores promediados. Esto último se realiza por estar ante unos pronósticos no bipartidistas, pues resulta lógico dar mayor importancia a los errores cometidos en los partidos más votados. Véanse Lewis (2005) y Hyndman y Koehler (2005).

Modelos

Para descubrir qué métodos de ponderación e imputación son los óptimos en los pronósticos electorales se han considerado dos resultados:

- El EMAP(E) obtenido independientemente para cada encuesta a partir de la intención de voto (o voto pasado en los estudios postelectorales) en las elecciones en consideración (modelos 1 a 4 de la tabla 1A del anexo).

Los predictores incluidos en estos modelos fueron:

- Año en el que tuvo lugar el proceso electoral (modelos 1 y 2).
- Clasificación de las elecciones según sean de cambio o continuidad (modelos 3 y 4).
- Tipo de encuesta (pre o postelectoral). Estas dos primeras variables sirven para controlar el efecto de la campaña y del clima político.
- Modalidades de ponderación (las siete expuestas en la tabla 2), a fin de comprobar las hipótesis 1, 2 y 4.

TABLA 3. Imputaciones incluidas en el diseño para recuerdo de voto e intención de voto

a) Recuerdo de voto (para ponderar por esta variable)		
Nombre*	Variable objetivo (modelo)	Variables predictoras
RV1 sin imputar		
RV2 imputado (NR)	Recuerdo de voto (multinomial)	Conjunto mixto **
RV3 imputado (NR y NVJ)	Recuerdo de voto (multinomial)	Conjunto mixto **
b) Intención de voto (para estimar su valor)		
Nombre	Variable objetivo (modelo)	Variables predictoras
1. No imputada		
2. Univariada básica	Intención de voto (multinomial)	Conjunto básico ***
3. Univariada ampliada	Intención de voto (multinomial)	Conjunto ampliado ****
4. Encadenada básica	Intención de voto (multinomial) Recuerdo de voto (multinomial) Ideología (ordinal)	Conjunto básico ***
5. Encadenada ampliada	Intención de voto (multinomial) Recuerdo de voto (multinomial) Ideología (ordinal)	Conjunto ampliado ****

* En el recuerdo de voto se imputó (método univariado) la no respuesta (NR) o a aquellos que no votaron en las anteriores elecciones por ser menores de edad (NVJ).

** Sexo, edad, tamaño de hábitat, comunidad autónoma, intención de voto e ideología.

*** Sexo, edad, nivel educativo, recuerdo de voto e ideología.

**** Sexo, edad, nivel educativo, recuerdo de voto, ideología, evaluación de la situación económica y evaluación de la situación política.

5) Modalidades de imputación de la intención de voto (las cinco expuestas en tabla 3.b⁷), para comprobar la *hipótesis 6*.

6) Interacción de forma de ponderación y año (modelos 1 y 2) o clasificación de las elecciones según sean de cambio o continuidad (modelos 3 y 4), contemplada en la *hipótesis 5*.

7) Interacción de modalidad de imputación y año (modelos 1 y 2) o clasificación de las elecciones según sean de cambio o continuidad (modelos 3 y 4), presente en la *hipótesis 7*.

b) El EMAP(R) del recuerdo de voto de las elecciones anteriores al sondeo en cuestión (modelo 5 de la tabla del anexo) para la prueba de la *hipótesis 3*. En este caso, los predictores fueron año, tipo de encuesta y modo de imputación del recuerdo de voto (los tres incluidos en la tabla 3.a).

Con los distintos tratamientos de imputación y estimación por elección y encuesta, se

⁷ Pero solo son posibles tres modalidades de imputación antes del año 2000 por no disponer en la encuesta de las preguntas de valoración del gobierno y la economía. A causa de ello, se desdoblan los modelos 1 y 3 (en los que se contemplan solo tres modalidades de imputación en todo el periodo analizado) en los modelos 2 y 4 en los que hay más modalidades de imputación, pero menos alcance temporal: solo las 6 últimas elecciones.

obtuvieron 595 estimaciones diferentes⁸ del EMAP(E) y 63 del EMAP(R).

Las hipótesis mencionadas fueron contrastadas en el interior de modelos de regresión, ajustados con mínimos cuadrados, mediante contrastes específicos de las medias estimadas de los errores medios absolutos ponderados. Una vez calculados los valores de F de estos contrastes, les fue aplicado el conservador ajuste de Bonferroni para evitar errores de tipo I (Rosenthal *et al.*, 2000).

RESULTADOS Y DISCUSIÓN

La primera hipótesis planteada en este trabajo versaba sobre la necesidad de emplear la ponderación sociodemográfica para mejorar la precisión de la estimación de voto. Decíamos que la mejora de la precisión tras equilibrar la muestra en términos de sexo, edad, estudios y actividad económica dependerá de si estas variables están relacionadas con la intención de voto y la probabilidad de responder a la encuesta (Särndal, 2007). El gráfico 1.a muestra que durante los años ochenta el uso de esta ponderación contribuía levemente a mejorar la estimación, aunque si se analiza el período completo no existen diferencias entre la estimación sin ponderar y la estimación ajustada para el perfil sociodemográfico ($F_{(1, 341)} = 2,59$; $p = 0,650$). El hecho de que las variables sociodemográficas no presenten una relación clara con la intención de voto o la propensión a responder no es sorprendente, ya que otros trabajos en los contextos americanos y británico apuntan en esa dirección (Crespi, 1988; Sturgis *et al.*, 2016).

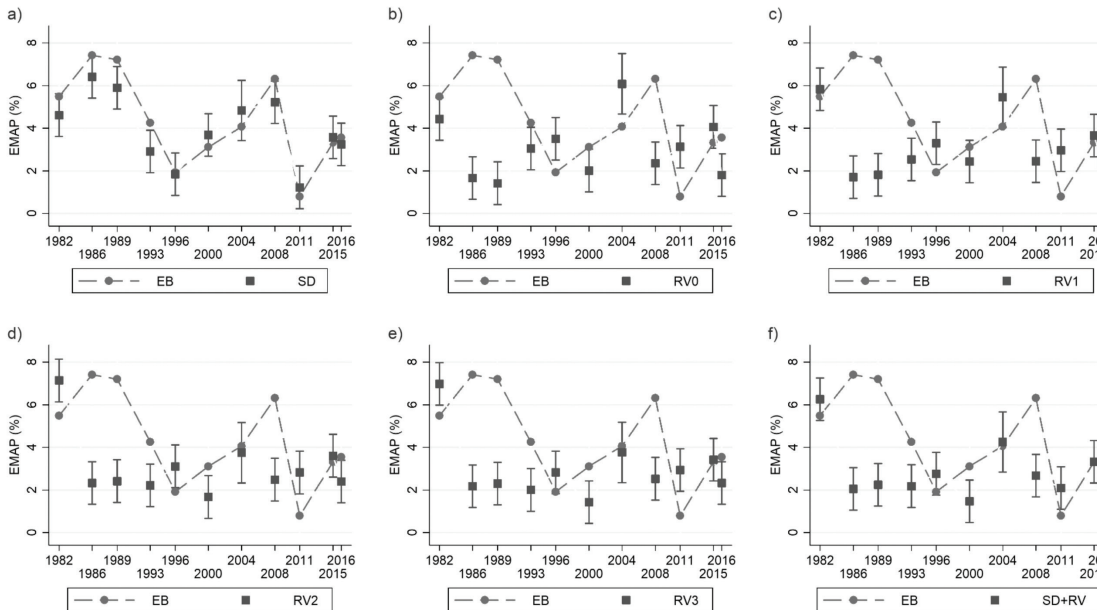
Ante la ineficiencia de utilizar factores sociodemográficos, una alternativa es acudir al recuerdo de voto, que está relacionado con la intención de votar y con la probabilidad de aceptar la visita de un entrevistador (Voogt y Saris, 2003). Los resultados muestran que, en general, el uso de esta variable en la ponderación ayuda a reducir el nivel de error que presentan las estimaciones de voto (gráficos 1.b-1.f).

Como se ha descrito en el apartado de la metodología, se ha trabajado con varias versiones de la ponderación por recuerdo de voto con el fin de averiguar si el tratamiento de esta variable mediante imputación repercute positivamente en la calidad de las estimaciones (hipótesis 2, 3 y 4). Dos de estas alternativas (RV0 y RV1) no emplean técnicas de imputación para corregir el posible sesgo derivado de la no respuesta parcial, mientras que las otras tres versiones sí hacen uso de esta técnica de corrección (RV2, RV3 y SD+RV).

Por lo general, aunque la información pública a este respecto es limitada, sabemos que las empresas demoscópicas hacen una transformación mínima de la variable recuerdo de voto antes de incluirla en la ponderación. En este trabajo hemos intentado replicar esa estrategia en dos ponderaciones. En el caso de RV0 se excluyen del proceso aquellos casos que no recuerdan o no contestan al recuerdo de voto, así como aquellos que manifiestan que no votaron en las anteriores elecciones. El uso de esta ponderación mejora, en el promedio de las elecciones estudiadas, la precisión de la estimación de voto, si se compara con el uso de la ponderación sociodemográfica ($F_{(1, 341)} = 15,98$; $p = 0,001$) o con la ausencia de ponderación ($F_{(1, 341)} = 31,44$; $p < 0,001$). El mismo comportamiento se observa al utilizar la ponderación RV1, que, al contrario que en el caso de RV0, incluye a aquellos que no desvelan qué hicieron en las elecciones anteriores o no votaron (sin ponderación: $F_{(1, 341)} = 27,22$; $p < 0,001$; sociodemográfica: $F_{(1, 341)} = 13,01$; $p = 0,002$). Este

⁸ Entre ellas, 441 no contienen imputaciones ampliadas y, por tanto, están disponibles en las 11 elecciones tratadas, mientras que 385 corresponden a las encuestas donde se pregunta por la valoración del gobierno y la economía y, por tanto, excluyen las predicciones anteriores al año 2000.

GRÁFICO 1. Comparación del Error Medio Absoluto Ponderado (EMAP) para la estimación de voto ponderada en sus diferentes versiones frente a la estimación base (EB) sin ponderar



La base de la comparación es la estimación base (EB) sin calibrar.

SD: Sociodemográfica (sexo, edad, región, actividad económica y estudios); RV0: Recuerdo de voto filtrando NS/NC y no votaron; RV1: Recuerdo de voto sin filtrar; RV2: Recuerdo de voto NS y NC imputados; RV3: Recuerdo de voto NS, NC y no tenían edad imputados; SD+RV: Sociodemográfica más RV3.

efecto global positivo contrasta con las retenciones mostradas por Worcester (1996) o las evidencias presentadas por Durand *et al.* (2015) para los casos francés y canadiense. A pesar de las deficiencias que pueda presentar esta variable, el balance global de su uso es positivo en el caso español para el período estudiado (hipótesis 2).

Una posible mejora de la ponderación por recuerdo de voto en las elecciones anteriores consiste en corregir las desviaciones de esta variable imputando valores válidos a aquellos que no contestaron. Las versiones RV2, RV3 (hipótesis 3) y SD+RV (hipótesis 4) son variantes de la ponderación por recuerdo de voto, en las que esta variable ha sido tratada previamente con técnicas de imputación múltiple. En el caso de RV2 se asignó un valor válido a aquellos que no respondieron, y en RV3, además, se imputó un valor válido a aquellos que en las anteriores elecciones

no tenían edad para votar. En ambos casos (gráficos 1.c y 1.d), el uso de la variable recuerdo de voto incrementa la precisión de las estimaciones en comparación con la ausencia de ponderación (RV2: $F_{(1, 341)} = 29,35$; $p < 0,001$; RV3: $F_{(1, 341)} = 34,95$; $p < 0,001$), o el uso de la ponderación sociodemográfica (RV2: $F_{(1, 341)} = 14,50$; $p = 0,001$; RV3: $F_{(1, 341)} = 18,51$; $p < 0,001$), aunque no existe una mejoría significativa si se comparan con las versiones del recuerdo de voto sin imputar (RV0 y RV1). En el caso de la ponderación SD+RV, que combina las variables sociodemográficas y el recuerdo de voto imputado (RV3), tampoco se obtienen mejores resultados que con el resto de las ponderaciones por recuerdo de voto, presumiblemente debido al efecto nulo de las variables sociodemográficas en la mayoría de las elecciones.

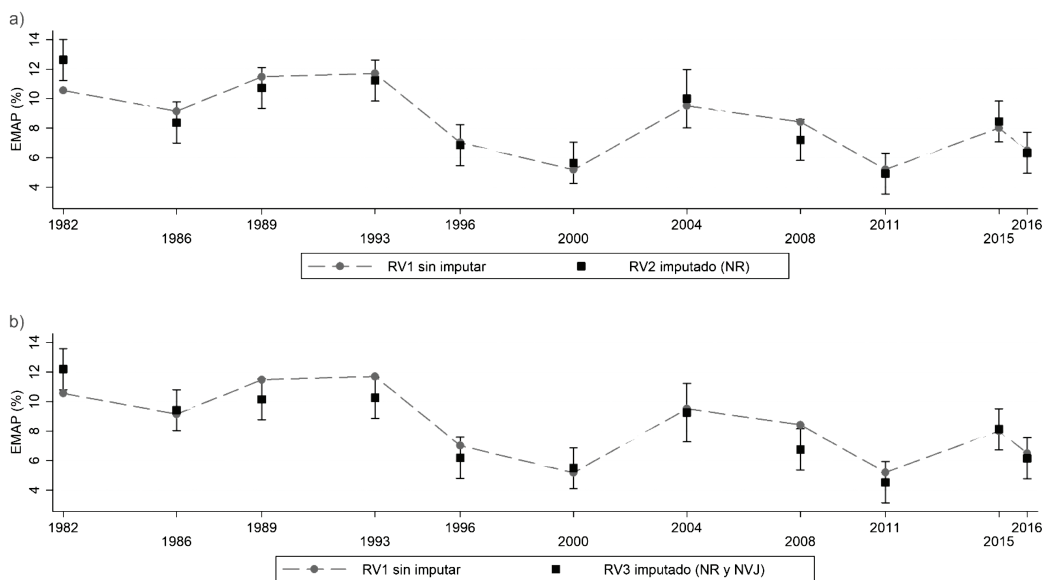
El uso de la imputación para atenuar el sesgo de no respuesta parcial en la variable

recuerdo de voto no supone una mejora en la precisión de las estimaciones ponderadas por esta variable. Este hecho se muestra en el gráfico 2: al ser pequeñas las correcciones aplicadas para asignar valores válidos a los que no contestan (RV2 imputado (NR)), o a los que no contestan y a aquellos que no pudieron votar en las elecciones anteriores (RV3 imputado (NR y ME)), apenas se reduce el nivel de error de la variable excepto en las elecciones de 1993 y 2008, donde hay mejoras de más de un punto porcentual. Esto puede deberse a dos factores: los predictores incluidos en la imputación y la magnitud del sesgo debido a la no respuesta parcial. Por un lado, los predictores incluidos en los modelos de imputación pueden no ser efectivos a la hora de discriminar entre las diferentes opciones del recuerdo de voto. Por otro lado, es factible que el sesgo generado por la no respuesta parcial sea mínimo y que, a pesar de su corrección, siga existiendo un nivel de error alto en el recuerdo de voto de-

bido a otros factores como la no respuesta total. Por ejemplo, esto último ocurriría si todos los que responden a la encuesta consiguen correctamente su comportamiento en las pasadas elecciones, pero son diferentes en su intención de voto que los que deciden no cooperar con el estudio.

Este efecto positivo del uso del recuerdo de voto en la ponderación, sin embargo, no es uniforme en el período estudiado (hipótesis 5). En línea con los hallazgos de Escobar *et al.* (2014), en las elecciones en las que se produce cambio en el partido de gobierno (1982, 1996, 2004 y 2011), el uso del recuerdo de voto no tiene efecto positivo sobre la estimación (p. ej., RV3: $F_{(1, 338)} = 5,39, p = 0,251$). Esta tendencia continúa en el nuevo ciclo político abierto en 2014, caracterizado por la irrupción de Podemos y Ciudadanos en el sistema de partidos (Orriols y Cordero, 2016; Rama, 2016), de forma que, en las elecciones de 2015, el uso del recuerdo de voto no contribuye a mejorar la estimación

GRÁFICO 2. Error Medio Absoluto Ponderado (EMAP) de la variable recuerdo de voto sin imputar (RV1) y las variables de recuerdo de voto imputadas (RV2 y RV3)



La base de la comparación es el recuerdo de voto sin imputar (RV1 sin imputar).
 RV2 imputado (NR): Imputado no respuesta (NS y NC); RV3 imputado (NR y NVJ): Imputado no respuesta y menores de edad (NVJ) en las últimas elecciones.

(p. ej., RV3: $F_{(1, 404)} = 0,01, p = 0,999$), lo que sí parece ocurrir en las encuestas de las elecciones de 2016, si bien el test de la comparación no arroja resultados contundentes (p. ej., RV3: $F_{(1, 404)} = 1,66, p = 0,198$).

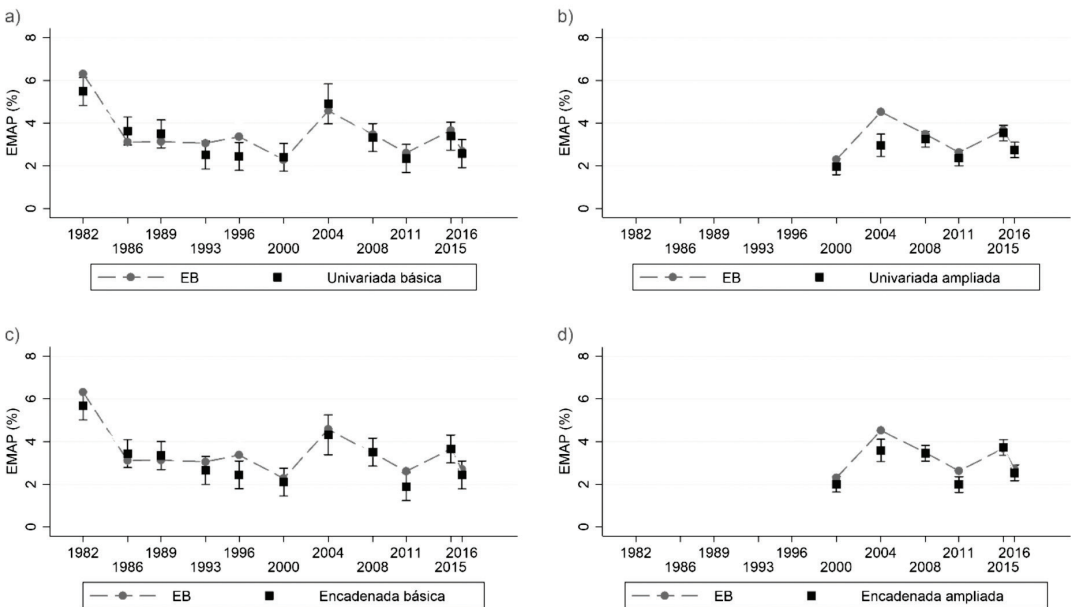
A la hora de estimar el voto, otro aspecto relevante son las posibles desviaciones de la intención de voto, debidas a que los que desvelan sus preferencias electorales tienen pensado votar de modo distinto a los que no las expresan. El uso de la imputación múltiple (hipótesis 6) puede servir para corregir estas posibles desviaciones (King, 2001; Liu, 2014). Para ello se han propuesto dos técnicas de imputación diferentes, la univariada y la encadenada, y dos conjuntos de variables, uno básico (sexo, edad, estudios, recuerdo de voto e ideología) y otro ampliado (las del conjunto básico más valoración del gobierno y de la situación económica).

La aplicación de la imputación univariada (gráfico 3.a) o encadenada (gráfico 3.c) empleando el conjunto de variables básico no

ayuda a mejorar las estimaciones. Solo en el caso de las versiones ampliadas (gráficos 3.b y 3.c), que incluyen la valoración que los encuestados hacen de la actuación del gobierno y de la situación económica —y que solo pueden ser computadas a partir de las elecciones del 2000—, se aprecia un tímido avance en la precisión de las estimaciones (simple: $F_{(1, 318)} = 12,97; p = 0,002$; encadenada: $F_{(1, 318)} = 8,51; p = 0,015$). En la misma línea, Liu (2014), que aplicó una metodología similar para corregir la intención de voto en las elecciones en Taiwán, muestra la nula capacidad de esta técnica para corregir posibles sesgos en la intención de voto. Sin embargo, con los datos dispuestos por el CIS, cabe colegir que la causa de la baja efectividad del modelo de imputación sea la inadecuada selección de predictores.

Por ello, en el caso de las versiones de la imputación ampliada se observa un efecto positivo, que parece reflejarse en las elecciones de 2004 y 2011. Sin embargo, no se puede demostrar que haya una tendencia clara según

GRÁFICO 3. Comparación del Error Medio Absoluto Ponderado (EMAP) para la estimación de voto imputada en sus diferentes versiones frente a la estimación base (EB) sin imputar



La base de la comparación es la estimación base (EB) sin imputar.

las elecciones sean o no de cambio (hipótesis 7), como ocurre en el caso de las ponderaciones. En parte esto ocurre porque la magnitud en que mejora la precisión de la estimación al utilizar la imputación ampliada es mínima.

CONCLUSIONES

La primera hipótesis que se planteaba en este trabajo trataba sobre la eficacia del uso de la ponderación sociodemográfica. Los resultados muestran que, a pesar de que hasta 1993 el uso de la ponderación produce una mejora limitada de las estimaciones, a partir de mediados de los noventa la contribución ha sido nula. Esto viene a coincidir con lo que ya se ha expuesto en otros estudios de caso (Crespi, 1988; Durand *et al.*, 2015). Ya sea porque las variables sociodemográficas están cada vez menos relacionadas con el voto o porque las muestras están relativamente equilibradas en este sentido, el efecto de esta ponderación sobre la estimación de voto es muy limitado. En línea con Sturgis *et al.* (2016), una de las claves está en emplear características poblacionales que sean capaces de corregir las desviaciones derivadas de la no respuesta.

En este trabajo, en consonancia con otros anteriores ya citados, se muestra el efecto positivo de ponderar por recuerdo de voto en el caso español. De esta manera, la segunda hipótesis queda confirmada. Sin embargo, aunque la muestra se equilibre con respecto a las preferencias de los electores en la última cita con las urnas, para que esta ponderación tenga éxito se precisa que los votantes de un partido que responden a la encuesta sean representativos de los que no participan. Este requisito nunca se cumple totalmente, y por ello la ponderación por recuerdo de voto aislada no es suficiente para eliminar por completo el sesgo presente en la estimación de voto.

Además, la propia variable recuerdo de voto puede presentar deficiencias, debido a

fallos de memoria de los entrevistados o a la deseabilidad social de la respuesta en determinados contextos (Crespi, 1988). Para corregir esas posibles deficiencias, en este trabajo se plantearon dos hipótesis, la primera es que la combinación de las variables sociodemográficas con el recuerdo de voto debería constituir la forma más efectiva de ponderar la estimación. La segunda se refiere a que la corrección de los posibles defectos de la variable recuerdo de voto podría ayudar a mejorar la calidad del ajuste. Ninguna de estas hipótesis se ha confirmado. Lo que sí se ha determinado es que la imputación múltiple, de la forma que se ha usado, ha servido para corregir solo levemente el sesgo presente en la variable recuerdo de voto. Este resultado, más que sembrar dudas sobre la imputación como técnica, resalta la importancia de los predictores seleccionados (Mercer *et al.*, 2018).

Por su parte, el uso de la imputación, como se planteaba en la cuarta hipótesis, tampoco ha resultado en una mejora sustancial de la precisión de las estimaciones. Una aportación significativa es que, en los modelos de imputación ampliados, en los que se añaden predictores sobre la valoración que hacen los entrevistados del gobierno y la situación económica, la capacidad de reducir el error en la estimación es mayor. Esto apunta, en línea con los hallazgos de Rivas *et al.* (2010), a que la elección de los predictores es la clave, lo que también es aplicable a la imputación del recuerdo de voto.

El efecto positivo de la ponderación, y de forma más tímida de la imputación ampliada, no es regular en el tiempo. Esta tendencia ya había sido detectada por parte de Escobar *et al.* (2014): en las elecciones de cambio político el efecto de la ponderación por recuerdo de voto es nulo y, en algunos casos, contraproducente. Esta tendencia se refleja también en este trabajo, al incluir los datos de 2015 y 2016. En las de 2015, que son consideradas elecciones de cambio político, el uso de la ponderación por recuerdo de voto tiene un efecto nulo, situación que se revierte parcial-

mente en 2016, elecciones en las que prácticamente se reproducen los resultados de 2015. En el caso de la imputación ampliada no se observa ninguna tendencia concluyente, aunque el efecto positivo se concentra en los estudios de 2004 y 2011, elecciones de cambio político en las que la ponderación por recuerdo de voto no funciona correctamente.

Convendría, en el futuro, dejar de lado las encuestas preelectorales por su escaso número de preguntas, y estudiar qué variables de las que están incluidas en los estudios postelectorales podrían arrojar mejores resultados en la mejora de la estimación de la conducta electoral de los ciudadanos mediante encuesta.

BIBLIOGRAFÍA

- Abrams, Mark (1970). «The Opinion Polls and the 1970 British General Election». *Public Opinion Quarterly*, 34(3): 317-324.
- Abramson, Paul R. (2007). «The French Presidential Election of 2007: Was Sarkozy the Condorcet Winner?». *French Politics* 5(3): 287-291.
- Anderson, Leslie (1992). «Surprises and Secrets: Lessons from the 1990 Nicaraguan Election». *Studies in Comparative International Development*, 27(3): 93-119.
- Ansolabehere, Stephen y Hersh, Eitan (2012). «Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate». *Political Analysis*, 20(4): 437-459.
- Bernhagen, Patrick y Marsh, Michael (2007). «The Partisan Effects of Low Turnout: Analyzing Vote Abstention as a Missing Data Problem». *Electoral Studies*, 26(3): 548-560.
- Bethlehem, Jelke; Cobben, Fannie y Schouten, Barry (2011). *Handbook of Nonresponse in Household Surveys*. New York: John Wiley & Sons.
- Biemer, Paul (2010). «Total Survey Error: Design, Implementation, and Evaluation». *Public Opinion Quarterly*, 74(5): 817-848.
- Biemer, Paul y Lyberg, Lars E. (2003). *Introduction to Survey Quality*. New York: John Wiley & Sons.
- Bosch, Agustí y Riba, Clara (2005). «Coyuntura económica y voto en España». *Revista de Sociología*, 75: 117-140.
- Caballé, Adriá; Grima, Pere y Marco-Almagro, Lluís (2013). «¿Aciertan los sondeos electorales? Análisis sobre la bondad de predicción de los sondeos electorales publicados en la prensa». *Revista Española de Investigaciones Sociológicas*, 143: 25-46.
- Callegaro, Mario y Gasperoni, Giancarlo (2008). «Accuracy of Pre-Election Polls for the 2006 Italian Parliamentary Election: Too Close to Call». *International Journal of Public Opinion Research*, 20(2): 148-170.
- Crespi, Irving (1988). *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation.
- Crewe, Ivor (1997). The Polls: Confidence Restored?. *Parliamentary Affairs*, 50: 569-585.
- Crewe, Ivor (2001). «The Opinion Polls: Still Biased to Labour». *Parliamentary Affairs*, 54(4): 650-665.
- Curtice, John (1997). *So How Well Did They Do? The Polls in the 1997 Election*. London: Centre for Research into Elections and Social Trends.
- D'Souza, John (2010). *Calibrate: A Stata Program for Calibration Weighting*. London: Stata User Group.
- D'Souza, John (2011). *Calibrate: Stata module to calibrate survey datasets to population totals, Statistical Software Components S457240*. Boston College Department of Economics.
- Durand, Claire (2008). «The Polls of the 2007 French Presidential Campaign: Were Lessons Learned from the 2002 Catastrophe?». *International Journal of Public Opinion Research*, 20(3): 275-298.
- Durand, Claire; Blais, André y Larochelle, Mylène (2004). «The Polls - Review. The Polls in the 2002 French Presidential Election: An Autopsy». *Public Opinion Quarterly*, 68(4): 602-622.
- Durand, Claire; Blais, André y Vachon, Sébastien (2001). «A Late Campaign Swing or a Failure of the Polls? The Case of the 1998 Quebec Election». *Public Opinion Quarterly*, 65(1): 108-123.
- Durand, Claire; Deslauriers, Melanie y Vallois, Isabelle (2015). «Should Recall of Previous Votes Be Used to Adjust Estimates of Voting Intention?». *Survey Methods: Insights from the Field*, 1-14. Disponible en: <http://surveyinsights.org/?p=3543>, acceso el 19 septiembre 2017.
- Escobar, Modesto y Jaime, Antonio M. (2013). «Métodos de imputación múltiple para predecir resultados electorales». En: Mendoza Velázquez, A. (ed.). *Aplicaciones en Economía y Ciencias Sociales con Stata*. Texas: Stata Press.

- Escobar, Modesto; Rivière, Jaime y Cilleros, Roberto (2014). *Los pronósticos electorales con encuestas: elecciones generales en España (1979-2011)*. Madrid: Centro de Investigaciones Sociológicas.
- Fuente, Ángel de la y Domenech Vilarino, Rafael (2015). *El nivel educativo de la población en España y sus regiones: 1960-2011*. BBVA Bank, Economic Research Department. Madrid: BBVA Research. Disponible en: https://www.bbva-research.com/wp-content/uploads/2015/02/WP_15-07_Educacion.pdf, acceso el 19 septiembre 2017.
- Fuente, Ángel de la (2015). *Series enlazadas de los principales agregados nacionales de la EPA, 1964-2014*. Instituto de Análisis Económico (CSIC). Madrid: FEDEA Research. Disponible en: <http://documentos.fedea.net/pubs/eee/eee2015-07.pdf>, acceso el 19 de septiembre 2017.
- Groves, Robert M. et al. (2013). *Survey Methodology*. New York: Wiley.
- Hyndman, Rob J. y Koehler, Anne B. (2005). «Another Look at Measures of Forecast Accuracy». *Monash Econometrics and Business Statistics Working Papers 13/05*. Monash University, Department of Econometrics and Business Statistics.
- Jaime, Antonio M. y Sáez Lozano, José L. (2001). *El comportamiento electoral en la democracia española*. Madrid: Centro de Estudios Políticos y Constitucionales.
- Jennings, Will y Wlezien, Christopher (2018). «Election Polling Errors across Time and Space». *Nature Human Behaviour*, 1.
- Jowell, Roger et al. (1993). «Review: The 1992 British Election: The Failure of the Polls». *Public Opinion Quarterly*, 57(2): 238-263.
- Katz, Daniel (1941). «The Public Opinion Polls and the 1940 Election». *Public Opinion Quarterly*, 5(1): 52-78.
- King, Gary et al. (2001). «Analyzing Incomplete Political Science Data». *American Political Science Review*, 85(1269): 49-69.
- Lago, Ignacio y Lago, Santiago (2005). «Does the Economy Matter? An Empirical Analysis of the Causal Chain Connecting the Economy and the Vote in Galicia». *Economics and Politics*, 17: 215-243.
- Lewis-Beck, M. S. (2005). «Election Forecasting: Principles and Practice». *The British Journal of Politics and International Relations*, 7: 145-164.
- Liu, Frank C. S. (2014). «Using Multiple Imputation for Vote Choice Data: A Comparison across Multiple Imputation Tools». *Open Journal of Political Science*, 4: 39-46.
- Lundström, Sixten y Särndal, Carl E. (2001). *Estimation in the Presence of Nonresponse and Frame Imperfection*. Sweden: Statistics Sweden.
- Lynn, Peter (2008). «The Problem of Nonresponse». En: European Association of Methodology (ed.). *The International Handbook of Survey Methodology*. New York: Lawrence Erlbaum Associates.
- Lynn, Peter y Jowell, Roger (1996). «How Might Opinion Polls Be Improved? The Case for Probability Sampling». *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159(1): 21-28.
- McEwen, Nicola (2004). «Opinion Polling in Scotland: An Analysis of the 2003 Scottish Parliament Election». *Journal of Elections, Public Opinion and Parties*, 14(1): 171-190.
- Mercer, Andrew; Lau, Arnold y Courtney, Kennedy (2018). *For Weighting Online Opt-In Samples, What Matters Most?* Washington: Pew Research Centre.
- Orriols, Lluís y Cordero, Guillermo (2016). «The Breakdown of the Spanish Two-Party System: The Upsurge of Podemos and Ciudadanos in the 2015 General Election». *South European Society and Politics*, 21(4): 469-492.
- Pavía, José M.; Badal, Elena y García-Cárceles, Belén (2016). «Spanish Exit Polls. Sampling Error or Nonresponse Bias?». *Revista Internacional de Sociología*, 74(3): e043.
- Pavía, José M. y Larraz, Beatriz (2012). «Sesgo de no-respuesta y modelos de superpoblación en encuestas electorales». *Revista Española de Investigaciones Sociológicas*, 137: 121-150.
- Rama, José (2016). *Crisis económica y sistema de partidos: síntomas de cambio político en España*. Barcelona: Institut de Ciències Polítiques i Socials.
- Rivas, Cristina; Martínez Rosón, María del Mar y Galindo, Purificación (2010). «La imputación múltiple como alternativa al análisis de la no respuesta en la variable intención de voto». *Revista Española de Ciencia Política*, 22: 99-118.
- Rivero, Gonzalo (2011). *Análisis de datos incompletos en ciencias sociales*. Madrid: CIS.
- Rosenthal, Robert; Rosnow, Ralph L. y Donald B. Rubin (2000). *Contrasts and Effect Sizes in Be-*

- havioral Research: A Correlational Approach*. Cambridge: Cambridge University Press.
- Rubin, Donald B. (1987). *Multiple Imputation for Non-response in Surveys*. New York: John Wiley & Sons.
- Sanders, David (2003). «Pre-Election Polling in Britain, 1950-1997». *Electoral Studies*, 22(1): 1-20.
- Särndal, Carl E. (2007). «The Calibration Approach in Survey Theory and Practice». *Survey Methodology*, 33(2): 99-119.
- Särndal, Carl E. y Lundström, Sixten (2005). *Estimation in Surveys with Nonresponse*. London: John Wiley & Sons.
- Sauger, Nicolas (2008). «Assessing the Accuracy of Polls for the French Presidential Election: The 2007 Experience». *French Politics*, 6(2): 116-136.
- Sciarini, Pascal y Goldberg, Andreas C. (2016). «Turnout Bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting». *Journal of Survey Statistics and Methodology*, 4(1): 110-137.
- Shlapentokh, Vladimir (1994). «The Polls - a Review - the 1993 Russian Election Polls». *Public Opinion Quarterly*, 58(46302): 579-602.
- Smith, Fred T. M. (1996). «Public Opinion Polls: The UK General Election, 1992». *Journal Of the Royal Statistical Society. Series A (Statistics in Society)*, 159(3): 535-545.
- Stata (2015). *Stata 14 Base Reference Manual*. College Station. Texas: Stata Press.
- StataCorp (2017). *Stata Multiple Imputation Reference Manual*. Texas: StataCorp.
- Sturgis, Patrick; Allum, Nick y Brunton-Smith, Ian (2009). «Attitudes Over Time: The Psychology of Panel Conditioning». En: Groves et al. (eds.). *Methodology of Longitudinal Surveys*. New York: Wiley.
- Sturgis, Patrick et al. (2016). *Report of the Inquiry into the 2015 British General Election Opinion Polls*. London: British Polling Council.
- Traugott, Michael W. (2005). «The Accuracy of the National Preelection Polls in the 2004 Presidential Election». *Public Opinion Quarterly*, 69(5 SPEC. ISS.): 642-654.
- Urquizu, Ignacio (2005). «El voto oculto en España». *Revista Española de Ciencia Política*, 13: 119-156.
- Varela, Jesús et al. (1998). «Estimación de la Respuesta de los 'No Sabe/No Contesta' en los Estudios de Intención de Voto». *Revista Española de Investigaciones Sociológicas*, 83: 269-287.
- Voogt, Robert J. J. y Saris, William E. (2003). «To Participate or not to Participate: The Link Between Survey Participation, Electoral Participation, and Political Interest». *Political Analysis*, 11(2): 164-179.
- Voss, Stephen; Gelman, Andrew y King, Gary (1995). «Preelection Survey Methodology: Details from Eight Polling Organizations, 1988 and 1992». *Public Opinion Quarterly*, 59: 98-132.
- Worcester, Robert (1996). «Political Polling: 95% Expertise and 5% Luck». *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(1): 5.

RECEPCIÓN: 25/09/2017

REVISIÓN: 14/02/2018

APROBACIÓN: 18/05/2018

ANEXO I. Modelos de regresión

TABLA 1A. Coeficientes de los modelos de regresión lineal (errores típicos entre paréntesis)

	M1	M2	M3	M4	M5
Elecciones					
1982 (base)					
1986	1,16 (0,81)				-1,41 (0,96)
1989	1,04 (0,81)				0,93 (0,96)
1993	-1,40 (0,81)				1,15 (0,96)
1996	-3,44*** (0,81)				-3,53*** (0,96)
2000	-2,84*** (0,81)				-5,38*** (0,96)
2004	-1,93 (1,00)	1,04* (0,48)			-1,04 (1,18)
2008	0,37 (0,81)	3,18*** (0,39)			-2,14* (0,96)
2011	-4,85*** (0,81)	-1,78*** (0,39)			-5,36*** (0,96)
2015	-2,56** (0,81)	0,53 (0,39)			-2,55* (0,96)
2016	-2,29** (0,81)	0,73 (0,39)			-4,06*** (0,96)
Cambio político					
Elecciones de continuidad (base)					
Elecciones de cambio			-1,97*** (0,47)	-1,62*** (0,36)	
Ponderaciones					
Estimación base (EB)					
SD	-0,87 (0,72)	0,56 (0,31)	-0,75 (0,39)	-0,27 (0,27)	
RV0	-1,06 (0,72)	-0,87** (0,31)	-3,26*** (0,39)	-2,09*** (0,27)	
RV1	0,34 (0,72)	-0,51 (0,31)	-3,10*** (0,39)	-1,74*** (0,27)	
RV2	1,65* (0,72)	-1,14*** (0,31)	-3,05*** (0,39)	-1,93*** (0,27)	
RV3	1,49* (0,72)	-1,33*** (0,31)	-3,18*** (0,39)	-2,00*** (0,27)	
SD + RV	0,77 (0,72)	-1,31*** (0,31)	-3,15*** (0,39)	-1,94*** (0,27)	

TABLA 1A. Modelos de regresión lineal (continuación)

	M1	M2	M3	M4	M5
Métodos de imputación					
Estimación base (EB)					
Univariada básica	-0,82 (0,47)	0,11 (0,26)	0,03 (0,25)	-0,05 (0,23)	
Univariada ampliada		-0,34 (0,26)		-0,18 (0,23)	
Encadenada básica	-0,63 (0,47)	-0,17 (0,26)		-0,13 (0,23)	
Encadenada ampliada		-0,30 (0,26)		-0,17 (0,23)	
Tipo de estudio					
Preelectoral (base)	-0,53***	-0,67	-0,62***	-0,84***	0,44
Postelectoral	(0,12)	(0,07)	(0,16)	(0,11)	(0,25)
Tratamiento RV					
RV1 sin imputar (base)					
RV2 imputado (NR)					2,07* (0,96)
RV3 imputado (NR y NVJ)					1,64 (0,96)
Constante	6,23*** (0,58)	3,26*** (0,28)	5,61*** (0,32)	4,65*** (0,25)	10,34*** (0,69)
F	7,66	16,66	11,71	11,35	11,24
Grados de libertad	99	66	18	22	33
P-valor	0,000	0,000	0,000	0,000	0,000
R cuadrado	0,60	0,73	0,30	0,37	0,85
Casos	441	385	441	385	63

M1: Error Medio Absoluto Ponderado (EMAP) de la estimación de voto en el período 1982-2016 solo incluye métodos de imputación con las variables básicas.

M2: EMAP de la estimación de voto en el período 2000-2016 incluyendo todos los métodos de imputación.

M3: M1 con la variable Elecciones sustituida por un indicador de cambio político (elecciones de cambio: 1982, 1996, 2004, 2011 y 2015).

M4: M2 con la variable Elecciones sustituida por un indicador de cambio político.

M5: EMAP de la variable recuerdo de voto en el período 1982-2016.

Ponderaciones. SD: variables sociodemográficas; RVO: recuerdo de voto filtrando NS/NC y no votaron;

RV1: recuerdo de voto sin filtrar; RV2: por recuerdo de voto con NS y NC imputados;

RV3 por recuerdo de voto NS, NC y no tenían edad imputados; RV+SD: por variables sociodemográficas y RV3.

Tratamiento RV (recuerdo de voto): RV1 sin imputar; RV2 imputado (NR): recuerdo de voto imputado a los que no responden;

RV3 imputado (NR y NVJ): recuerdo de voto imputado a los que no responden y no tenían edad para votar.

Las interacciones han sido omitidas para facilitar la interpretación.

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,01$.