Making sense of pollsters' errors. An analysis of the 2014 second-order European election predictions

Marco Giuliani[a]

[a]*Department of Social and Political Sciences, Università degli studi di Milano, Milano, Italy*

Via Conservatorio 7, 20122 Milano, Italy; marco.giuliani@unimi.it

ORCiD: 0000-0002-6927-7177

PhD in Political Science (1990) from the University of Florence (Italy), full professor at the University of Milano where he teaches Comparative politics. His main research topics are the comparative analysis of the performance of democracies, electoral behavior and EU politics and policy-making.

Abstract

Pollsters have been recently accused of delivering poor electoral predictions. We argue that one of the reasons for their failures lies in the difficulty of including an updated deep understanding of electoral behaviour. Even if pollsters' predictions are not forecasts produced by models, the set of choices needed to produce their estimates is not indifferent to a theoretical comprehension of electoral dynamics. We exemplify this lack of theory by using an original dataset consisting of 1057 party*poll observations in the case of the last European election. Pollsters failed to account for what we know about second-order elections, thus overestimating government and big parties, which normally obtain poor results in European elections, and underestimating new and Eurosceptic ones, which usually perform well.

Keywords: Election polls; Accuracy; Second-order elections; European politics

Introduction

Pollsters have had some hard times lately. It seems that the more their work becomes influential, even endogenously affecting what they are trying to predict, the less it is reliable. They did not anticipate very diverse outcomes, such as Trump's US presidency, Brexit, Fillon winning the primaries for the Republican candidate in France, the Conservative victory in the UK 2015 general election, the success of Likud in Israel in that same year, or the partial stalemates produced in the Italian 2013 election and in the German 2017 one. And the list could easily continue. Their failures are often attributed to a set of contingencies, like a "perfect storm" conspiring against their prediction capabilities.[1] But there is probably a different set of factors involved.

Scholars have identified the technical elements that affect pollsters' predictions and make their work complex (Voss, Gelman and King 1995). Pooling the polls (Jackman 2005) is often seen as a remedy against potential biases due to house effects, and a strategy for tracking actual changes in voter support. However, it would not have been enough in the cases cited above. Deriving predictions from pre-election surveys is not a straightforward task. A series of corrections and adjustments are needed in order to produce a plausible guess from a sample which is probably non-representative of relevant characteristics of the population, not least the willingness to respond to surveys. Post-stratification weights are applied to correct for that well-known problem. "Pollsters also employ weights to adjust for the likelihood that respondents will cast a ballot, the differential response to polls between supporters of different parties, and their differential willingness to admit preferences to the pollsters. (…) Besides weighting, there are other practices that can alter error variance, (…and) lead to a net reduction or net increase in error variance compared to a purely random sample" (Fisher et al. 2011, 251).

A set of implicit assumptions and consolidated knowledge is crucial in order to apply those corrections and adjustments. This wisdom pertains to traditional electoral behaviour, local dynamics, general trends, etc. It is mostly derived from previous experiences, and should be constantly updated. Yet, if society moves through troubled and unparalleled circumstances – a deep economic crisis, a series of external shocks, new world dynamics – the past may be only marginally useful for predicting the future. In that event, a deeper theoretical understanding is needed in order to realign that wisdom with the new shaken practices. We argue that this is precisely what the above-cited pollsters' failures have in common. After a prolonged period of social unrest – due to the Great Recession, international tensions, migration and globalization dynamics – social scientists need to go back to some deeper theoretical understanding of the dynamics investigated in order to regain the capacity to adjust their predictions positively. It is probably not by chance that political scientists' forecasts were closer to the actual results of the US Presidential competition than pollsters' predictions (Campbell 2016).[2]

In this article, we test this intuition by using ex-post the 2014 European elections as a case study. Its advantage is that of having parallel competitions in 28 member states, with very different polling firms making projections on a wide range of diverse parties. Because of the complexity of multiple multi-party ballots, it is a crucial case study, with the further benefit of having a well-established interpretation of voters' behaviours in the theory of second-order elections. We argue that failing to account for that theoretical understanding was a systematic part of the misfit between predictions and actual results, which, though not as dense with consequences as in the cases cited, was not trivial, with an average discrepancy of more than 2%. The systematic bias in the predictions that we ascertain in what follows is thus the outcome of mechanically

applying long-established polling practices developed in several rounds of national general elections; practices that are not (or no longer) appropriate to this specific electoral context. If voting behaviour is at least partially context-dependent, with citizens making up their minds in different moments and following different frames, the same should be said of the polling practices and adjustments mentioned before by Fisher and colleagues.

The article is organized as follows. In the following section, we present the theory of second-order elections applied to European ballots. From that theory we derive five hypotheses, which represent the symmetrical expectations of pollsters' errors regarding the electoral performances of different types of party. Next, we introduce our original dataset collecting the predictions of several different polling firms in the 28 EU member states in the event of the May 2014 common European election. We further specify the independent and control variables that will be used in our econometric models, and, for robustness, propose several different dependent variables measuring the gap between predictions and actual results. Finally, we present the results of those models, which will be further summarized and discussed in the concluding section.

Theory and Hypotheses

Scholars of European politics generally agree that elections for the European parliament should be considered second-order elections (Marsh 1998, Reif and Schmitt 1980). Citizens generally perceive them as less important than national general elections because there is less at stake. Most importantly, notwithstanding the institutional innovations introduced by the Lisbon Treaty, their results are not directly connected to the investiture of a common European executive. In spite of the evolution of the EU, and of the politicization of that governance level, most scholars agree that even the most

recent elections adhere to the second-order framework (e.g. Hix and March 2011; Schmitt and Toygür 2016). As a consequence, voters may consider behaving in the European polls differently from their habits during their national general elections.

First of all, they may simply decide not to vote. For this reason, the level of turnout is generally lower in European elections. In the last 2014 European election, the turnout in all member states except Belgium – which simultaneously ran its national and local election – was lower than in their previous general elections. On average, the difference was almost 25% of the voting age population. This probably does not significantly affect the legitimization of the European parliament (van der Eijk and Schmitt 2009), which in many countries is trusted more than the national counterpart, but disturbs the work of pollsters for three different interconnected reasons. Firstly, because it is more difficult to guess the behaviour of a smaller non-random sample of the population, which is different from the one more usually surveyed for national elections. Secondly, because turnout may affect the electorates of the diverse parties in different ways, not necessarily coincident with what happens in domestic elections. Lastly, because we do not know much about the uneven disposition of those electorates to declare sincerely their lack of interest in participating in second-order elections.

This strand of literature suggests a second behavioural option which is a direct consequence of the less-at-stake interpretation of second-order elections: that of voting more sincerely. This conduct is normally considered favoured also by institutional elements, such as the specific design of the electoral system which features a proportional formula with comparatively large districts, as is confirmed by comparing the so-called ‘effective threshold’ in the two types of elections. This set-up relatively favours, or does not discourage, the performance of small, and even new parties. However, indices synthesising the disproportionality of the system, normally associated

with strategic behaviours, do not entirely confirm that interpretation. The average Gallagher index for European and national elections in our sample is not systematically different, with some countries having a higher and others a lower apparent level of permissiveness. Yet, the actual value for the European parliament could have been endogenously produced by the increased competition amongst a larger number of political groups, as demonstrated by the fact that the effective number of electoral parties is substantially higher, almost 1 point, for the European election compared to the national one, something that indirectly confirms the favourable institutional environment of European elections for small and new parties.

Moreover, there are further, political more than institutional, reasons why we should observe a greater success of new or usually marginalized parties during European elections. The literature normally connects them with the specular suffering of government parties: on the one hand, because the latter cannot appeal to any sense of responsibility, or to the risk of wasting votes in those electoral appointments, as they can profitably do in the national arena; on the other, because voters may profit from the opportunity to experiment with a different vote. This means that citizens choose parties closer to their preferences, yet even without any domestic coalitional potential, and they empower leaders, political groups and MEPs whose actions and choices they will evaluate in the less sensitive arena of Strasbourg and Brussels.

There is even a third argument that has been put forward to expect a reduced consensus for governing parties during European elections. Several empirical studies, in very different institutional contexts, have found evidence of a recurrent electoral cycle for incumbents. After a period of honeymoon, they usually lose consent, at least partially recovering it on the eve of a new election (Fisher 2014). This may depend on the citizens' strategic expression of discontent for the government performance, as the

traditional second-order theory goes, or be the side-product of a lack of information and mobilization from behalf of incumbents that are mainly interested in the first-order arena (Weber 2007).

Whatever the underlying mechanism, and unless a member state holds simultaneous elections or has had its preceding national ballot very close to the European one – something that happened in 2014 only for Belgium and Hungary – we should expect government parties to lose in what approximates a mid-term election (Reif 1984).

European elections, at least the most recent ones, have one further feature that distinguishes them from the several other national second-order consultations. Potentially, part of the competition takes place along a supranational dimension that confronts Europhiles, Eurosceptic and neutral positions. As a consequence, parties with a clearer profile and position on EU integration should perform better in those elections. We must accompany this statement with a word of caution, however, because one of the major failures of the EU has been precisely its incapacity to foster any sincere debate on European issues. Yet some recent studies have demonstrated that information makes a difference between those who vote only according to domestic preferences, and those who also consider issues of European integration (Hobolt and Wittrock 2011). Furthermore, at least for 2014, the austerity policies of the EU against the Great Recession undoubtedly affected the electoral debate in many countries (Schmitt and Teperoglou 2015). This may have produced a stronger polarization of the electorate on this, now more salient, dimension. It is a political dynamic that, due to the economic circumstances of the Great Recession, should have favoured mostly Eurosceptic parties (Giuliani and Massari 2017), with the UKIP being the clearest example of this tendency.

To synthesise, the literature on second-order elections suggests that in the European polls, compared to national appointments: a) government parties lose, especially if elections fall around mid-term; b) there is relatively more space for small, and even new parties; c) parties with an unambiguous position on the EU, which in 2014 mostly meant Eurosceptic parties, comparatively win. Admittedly, these features are not independent. Incumbent parties, for example, cannot be new (unless they formed by fission or aggregation during the ongoing legislature), are mostly large in size, and with a comparatively moderate if not positive attitude towards Europe. If our aim here were to explain the level of parties' support in European elections, as Hix and Marsh (2011) did for the first seven appointments, and Schmitt and Toygür (2016) for the last one, we would certainly need a multivariate model including all those variables at the same time. Instead, we are here only exploring potential overlapping sources of pollsters' bias, and will thus mostly check them separately, following what also the original literature on second-order elections did.

Therefore, if our intuition is right, and pollsters, accustomed to predicting electoral results within a domestic frame, failed to take proper consideration of the theoretical and empirical suggestions specific to second-order European elections, we should observe an overestimation of parties that the literature expects to lose, and an underestimation of those that it expects to win. Thus:

- Hp. 1 Pollsters overestimated government parties,
- Hp. 2 … mostly if the European election fell in the middle of the electoral cycle;
- Hp. 3 Pollsters overestimated big parties,
- Hp. 4 … and underestimated new ones;
- Hp. 5 Pollsters underestimated Eurosceptic parties.

Data, Variables and Measures

In order to test our hypotheses, we collected all the final electoral predictions formulated by different polling firms in each EU member state before the European election held from 22 to 25 May 2014.[3] In some cases, polls were published just before the election, with 45% of them made public in the last 10 days before the electoral appointment, and almost all of them during the last month. Eventually, we arrived at 1057 party*poll observations for which we had both a prediction and an actual result.

Mainly from those two pieces of information, we computed several different measures of our dependent variable, that is, the under- or over-estimation of a party's performance. The simplest of these indices is the *Error* of the prediction, i.e. the difference between the estimate and the actual result, which assumes positive values in the case of an overrated party, and negative ones in the case of an underrated one. Since our hypotheses deal with the direction of the blunder, more than with its magnitude, we also computed a dummy variable, *Positive error*, when the sign of the error reflected overestimation of the party.[4]

The statistical soundness of each prediction was not our major concern, since we were mostly interested in detecting potential cross-cutting inaccuracies. Therefore, even systematic small differences between estimate and actual result, within the usual sampling margin of error, contributed to, and were included in, our tests. However, we decided to replicate our analyses by also adopting a more conservative approach. Using the sample size to estimate the 95% confidence intervals of the estimated result of each party, we verified when our error exceeded that random component, thus denoting a biased prediction. We thus built two non-symmetric dummy variables, *Positive bias*

and *Negative bias*, assuming the value of 1, respectively, in the case of non-random positive and negative prediction mistakes, and zero in all the other circumstances.

Finally, we took up a recent suggestion by Arzheimer and Evans (2014) in regard to measuring the polling bias, which generalizes to a multi-party competition an index originally proposed by Martin, Traugott and Kennedy (2005) for two-party races. Their *Accuracy* index is the following:

$$A_i' = ln\left(\frac{p_i}{1-p_i} \times \frac{1-v_i}{v_i}\right) \qquad (1)$$

where p_i and v_i represent respectively the proportion of the votes predicted by the polls, and those actually obtained in the election by party *i*. "Positive values indicate bias in favour of party *i*, whereas negative values imply bias against *i*" (33). Adopting several measures for the same concept representing our dependent variable served as a robustness check of our results.

Our independent variables are easily defined. *Government* and *New* parties were measured as dummy variables relative to the incumbent executive, and to the preceding national ballot. In most cases, the correspondence between the domestic and European arenas was straightforward. In case of doubt, we mainly preferred to resolve it in favour of novelty, thus producing a more conservative test for our hypotheses.[5] *Euroscepticism* is simply the reversed scale of the party positions on EU integration proposed by Chapel Hills experts for 2014 (Bakker et al. 2015), here ranging from 1 (Strongly in favour) to 7 (Strongly opposed), only marginally complemented with qualitative information in less than 3% of our observations, for minor parties excluded by their analysis. The *Size* of the party was computed as the percentage of votes obtained in the preceding general election, and the electoral *Cycle* as the ratio between the number of days since that domestic appointment and the overall legal duration of the parliamentary term.

We further introduced a small set of standard control variables capturing typical problems of electoral polling and predictions; and because of some missing information on these controls the actual number of observations used in our models was reduced to approximately 930 cases. First of all, we included in the right-hand side of the equation the size of the *Sample*, and the *Time* between the date of the poll and the European election computed in number of days (Jennings and Wlezien 2016). Next, we controlled for the level of *Turnout*, which, as we have seen, is one of the challenging elements of European elections, strongly affecting our prediction capacity. Finally, we checked the *Contemporaneous* holding of other elections or referenda, and the presence of *Compulsory* voting, two factors that may reduce the context-specific characteristic of the European appointment.

Models and Empirical Results

Before testing our hypotheses thoroughly, we simply verified the plausibility of the effect of being in government and being a new party – the only two dummy independent variables involved in our conjectures – on pollsters' predictions. If they overlooked what we know about second-order elections, we should have observed the former overestimated, and the latter underestimated in their predictions. As can be seen in Table 1, the average difference between polls and results is in fact higher for government parties compared to opposition ones, and the accuracy index is positive in the former case and negative in the latter.[6] The Anova tests confirmed that the differences between the two values is always significant at a $p<0.05$ value. The same applies to the comparison between the mean values for new and old parties, although this time only the difference between values of the error was statistically significant, whereas the accuracy index failed to reach the usual standard. Given the promising results, and the tiny differences, we proceeded with more robust confirmations

including our control variables, and the whole range of indices presented in the previous section.

Table 1. Comparison of Error and Accuracy of the prediction for different party characteristics

		Error		Accuracy	
	N	Mean	Std. err.	Mean	Std. err.
Government	291	0.81	0.27	0.08	0.02
Opposition	766	0.10	0.10	-0.03	0.01
New	182	-0.05	0.16	-0.03	0.04
Old	866	0.37	0.12	0.01	0.01

We started by modelling the three dichotomic outcomes of having overestimated predictions (looking for the causes of (a) positive errors and (b) positive biases) or underestimated ones (explaining the origins of (c) negative biases). For this purpose, we used separate logistic regressions, clustering the standard errors for each poll in order to account for the non-independence of the observations within the same wave.[7] Given our effects-of-causes research design (Mahoney and Goertz 2006), only marginally interested in the comprehensive explanation of the dependent variable, we only report the coefficients for the covariates of interest of the three models, referring to the online appendix for the complete results.

Table 2. Impact of different party characteristics on its over and under-estimation in the polls

	Government	Size	New	Euroscepticism
Positive error	0.317**	0.042***	-0.417**	-0.031
	(0.143)	(0.007)	(0.186)	(0.032)
Positive bias	0.679***	0.048***	-0.265	-0.171***
	(0.163)	(0.008)	(0.242)	(0.042)
Negative bias	0.019	-0.031***	0.175	-0.065
	(0.162)	(0.009)	(0.227)	(0.044)

Constant and control variables not reported; Clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Confirming our expectations, Table 2 shows a systematic positive difference between predictions and results, with $p<0.05$ for government parties, and with $p<0.01$ for big ones. More specifically, when the odds ratio is computed from the coefficients reported in the table, being in government increases the probability of being overestimated by pollsters by more than 37%, while each increase of 1% in size boosts that same probability by more than 4%. At the same time, being a new party systematically reduces, by 32%, the chances of being overrated, while the degree of Euroscepticism does not seem to affect that result.

If we limit the positive outcomes to overestimations beyond the sampling error, thus looking at positive biases, the results are mostly the same. Once again, the coefficients of the models for being in government and for being a large party are both positive and highly statistically significant. The probability of a positive bias is almost double for incumbents compared to non-incumbents, and grows by 5% for each increase of 1% in size. This time, the coefficient for being a new party is non-significant, though with the correct sign[8], whereas the position on the EU integration process is, as expected, inversely associated with overestimation. For each point increase on the

seven-point scale of Euroscepticism there is an almost 16% decrease in the probability of a positive bias in pollsters' estimates. In the models explaining negative biases, which are not simply the opposite of positive ones because of the grey area of errors within the sampling margins, only the size of the party is statistically significant. The coefficient is negative, as expected, since the higher the support in the preceding general election, the lower the probability of being underestimated by pollsters.

Our hypotheses are largely confirmed by these first results, though it should be noted that, in the last two models, the non-symmetric nature of bias reveals that a systematic reduction of the probability of being overrated does not translate into a specular increase in the chances of being underestimated, and vice versa. For symmetric, and even more fine-grained hypotheses testing, we should turn to our last two dependent variables, which consider not only the direction of the misjudgement but also its extent. For this reason, using the same control variables and clustering the standard errors as before, we ran a series of OLS regressions aimed at explaining the gap between predictions and actual results, and the proposed index of accuracy. The results are presented in Table 3.

Table 3. Impact of different party characteristics on the accuracy and on the error of the party prediction

	Government	Size	New	Euroscepticism
Error	0.551*	0.075***	-0.292	-0.032
	(0.286)	(0.012)	(0.242)	(0.052)
Accuracy	0.093***	0.005***	-0.031	-0.014*
	(0.027)	(0.001)	(0.055)	(0.007)

Constant and control variables not reported; Clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

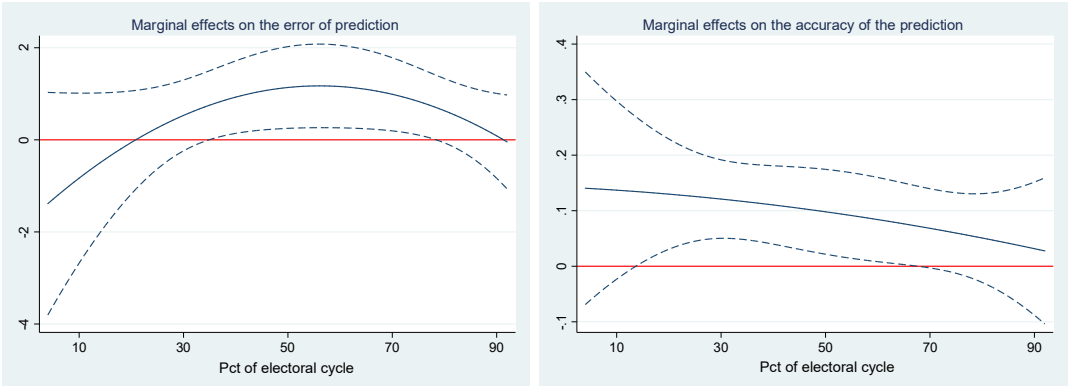
These analyses confirm the picture that emerged from the logit models. All the

coefficients have the expected signs, and in five cases out of eight they are statistically significant. Euroscepticism is the only exception if we take the error between polls and actual results as dependent variable, whereas being a new party systematically affects neither the error nor the accuracy of the prediction. Turning to the positive evidence, being incumbent yields on average a surplus estimate of half a point, while for every increase of 10% in the votes obtained in the previous national election there is a ¾ point of overestimation in the European one. Turning to the index of accuracy, which has a 4-point empirical range, almost one tenth of a point of favouritism is generated by incumbency, or by a party holding 25% of the votes at the national level, whereas moving from the lowest extreme of the scale of Euroscepticism to the highest one produces a discrimination of a similar magnitude.

Generally speaking, combining different econometric models with four ways of measuring our dependent variable – the misjudgement of pollsters – yields not a perfect but a sufficiently robust confirmation of our hypotheses. Big and incumbent parties obtain favourable estimates, while new and Eurosceptic ones receive mostly adverse predictions. There is only one of the proposed propositions that has not yet been tested and that cannot be investigated with our direct models: the one suggesting that government parties should be mostly overestimated the farther away they are from the preceding or the successive national election (because the second-order theory expect them to lose the most when the European appointment falls exactly in the middle of the mandate; e.g. Reif 1984; Weber 2011). In order to test this hypothesis, we ran two conditional models in which the electoral cycle interacted with our dummy variable capturing the characteristic of being incumbent parties. Following the best practices, more than reporting the coefficients, it was essential to plot the marginal conditional effects for the whole observed range of the electoral cycle (Kam and Franzese 2007,

Brambor, Clark and Golder 2006). Since our hypothesis assumed a non-linear relationship with the time from the preceding election, we modelled it using the square of the cycle variable as interaction term. The results are illustrated in Figure 1.

Figure 1. Marginal effects of incumbency on the error (left panel) and on the accuracy of the prediction (right panel) at different moments of the electoral cycle



Both graphs, representing the marginal impacts of incumbency on the delta between polls and results, and on the accuracy index, confirm our intuition and the expectations derived from the theory of second-order elections. In the first part of the electoral cycle, the honeymoon effect prevents government parties from suffering the usual impact of European appointments. Symmetrically, as shown by the confidence intervals overlapping the no-effect horizontal line, they are not overestimated by pollsters, whose predictions are made easier also by the closeness of the preceding electoral ballot. The same happens in the last phase of the electoral cycle, when government parties usually recover, and when, once again, pollsters' predictions become more accurate, cancelling any systematic bias in their favour. However, when European elections take place during the central phases of the cycle, approximately between one and three years after the preceding national appointment, government parties are more exposed to second-

order effects, and pollsters naively overestimate their actual support. This can be seen from the positive marginal effects shown in both graphs, whose confidence intervals are consistently above the horizontal zero line in the central portion of the chart.

The two graphs are far from being perfectly symmetrical, and especially the point estimates of the one regarding accuracy resemble a linearly declining over-prediction. This is probably due to the fact that two factors combine in producing those marginal effects: on the one side, second-order elections modify political behaviours, and, on the other, the closeness of other ballots provides additional information which is probably biased by the contingent character of that appointment. In any case, what matters more for our propositions, which reverse the predictions of the theory of second-order elections, is that confidence intervals overlap with the null hypothesis at the two extremes of the cycle, thus confirming also our final (second) hypothesis.

Conclusion

In this article, we have empirically checked one simple hypothesis: that pollsters systematically underutilize the theoretical knowledge produced by the social sciences, and by political scientists more specifically. By using the 2014 European election as a test bed for that insight, we have suggested that pollsters erred in the direction opposite to the expectations of the theory of second-order elections, which found confirmation even in that event. For robustness, we have tested several measures, mostly confirming our hypotheses.

When the theory predicts European success, as in the case of new and Eurosceptic parties, pollsters generally underestimated their electoral performances, thus confirming our hypotheses 4 and 5. When the theory suggests a relative European failure, as for big and government parties, pollsters mostly overestimated their actual

electoral backing, as we suggested with our hypothesis 1 and 3. Including a conditional effect of the electoral cycle better specifies the misfit between predictions and actual results due to a failure to consider the incumbency effect in second-order elections, as we actually implied with our hypothesis 2.

The modern world, and not only our political worlds, is becoming more and more complex and less and less predictable (Waldrop 1992; Bertuglia and Vaio 2005). That does not mean that, even in such complexity, equilibria do not form and regular patterns cannot be identified. Bringing this very general idea to the daily work of pollsters as social scientists, it is evident that the more complex and changing a social and political environment is, the less we can simply rely on a correct sampling design. Predicting electoral results "doesn't just involve asking people whether they support candidate A or candidate B. It also involves trying to determine whether respondents will act on their preferences by casting a ballot at all. […] It is this extra step, […] that is quite distinct from the principle of random sampling and good question design that make survey research valid and reliable" (Gramlich 2017).

The problem is thus no longer, or not simply, statistical, and even house effects are irrelevant, as we indirectly demonstrated with our cross-country and cross-polling analysis. Several correcting strategies are needed, including all the possibilities offered by modern data analytics. Yet, corrections due to firmly-established theories should be the first to be taken into consideration. For the European appointment polls, we used the theory of second-order election, which could be further complemented with more general hypotheses regarding the support cycle of government parties (Fisher 2014; Weber 2011) and the specific profile of late voters (Box-Steffensmeier et al 2015).

Going back to the examples mentioned in the introduction of this article, what could be other theories in social and political science that are sufficiently established to

provide useful guidance to pollsters? The first obvious candidate is the theory of economic voting, whose robustness has been proved in different contexts and time-periods (Lewis-Beck and Stegmaier 2007; Stegmaier, Lewis-Beck and Park 2017; Giuliani and Massari 2018). Its conjectures give an additional leverage to non-incumbent parties, clarify the electoral costs for coalition partners, and may complement other information regarding the expected turnout levels: three suggestions that certainly apply to many of the prediction failures mentioned above.[9]

The theory of valence political competition (Curini 2018) is a second contender for that role. The less ideological the competition, and the more issues such as corruption, leadership capacities, and similar non-positional topics appear at the forefront of the campaign, the more its hypotheses may help. At the intersection between political communication and political psychology, the same could be said about works on the personalization of electoral competition (Garzia 2014), on the self-reinforcing effects of echo-chambers (Barberà 2015) – whose effects are exploited by the analysis of social networks (Leiter et al. 2018) – and on the risks of spirals of silence (Noelle-Neumann 1993).

It should be remembered that we are thinking neither of substituting the mining of public opinion, nor moving directly from predictions to forecasts. These theories may help improve pollsters' work much in the same vein as more traditional sociological knowledge helped them frame the procedures necessary for post-stratification or for non-respondents substitution in the past. Even "local" theories accounting for regular behavioural patterns and voting traditions in certain countries may be of use, until our increasingly fluid and unpredictable world endangers even those solid theoretical anchors.

References

Arzheimer, K., and J. Evans. 2014. "A New Multinomial Accuracy Measure for Polling Bias." *Political Analysis* 22 (1): 31-44.

Bakker, R., E. Edwards, L. Hooghe, S. Jolly, G. Marks, J. Polk, J. Rovny, M. Steenbergen and M. Vachudova. 2015. *2014 Chapel Hill Expert Survey. Version 2015.1*. Available on chesdata.eu. Chapel Hill, NC: University of North Carolina.

Barberà, P. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23: 76–91.

Barnea, S. and G. Rahat. 2011. " 'Out with the old, in with the "new" ': What constitutes a new party?" *Party Politics* 17 (3): 303-320.

Bertuglia, C.S. and F. Vaio. 2005. *Nonlinearity, Chaos, and Complexity. The Dynamics of Natural and Social Systems*. Oxford: Oxford U.P.

Bolleyer, N. and E. Bytzek. 2013. "Origins of party formation and new party success in advanced democracies." *European Journal of Political Research* 52 (6): 773–796.

Box-Steffensmeier, J., M. Dillard, D. Kimball and W. Massengill. 2015. "The long and short of it: The unpredictability of late deciding voters." *Electoral studies* 39: 181-194.

Brambor, T., W. Clark and M. Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63-82.

Campbell, J. E. 2016. “How Accurate Were the Political Science Forecasts of the 2016 Presidential Election?.” *Cristal Ball*, University of Virginias Center for Politics, November 17.

Curini, L. 2018. *Corruption, Ideology, and Populism. The Rise of Valence Political Campaigning*. New York: Palgrave.

Döring, H., and P. Manow. 2015. *Parliaments and Governments Database (ParlGov): Information on Parties, Elections and Cabinets in Modern Democracies*. Available at http://www.parlgov.org/

Emanuele, V. and A. Chiaramonte. 2016. “A growing impact of new parties: Myth or reality? Party system innovation in Western Europe after 1945.” *Party Politics* DOI: 10.1177/1354068816678887.

Fisher, S. D., R. Ford, W. Jennings, M. Pickup, and C. Wlezien. 2011. “From Polls to Votes to Seats: Forecasting the 2010 British General Election.” *Electoral Studies* 30 (2): 250–257.

Fisher, S. D. 2014. “Predictable and Unpredictable Changes in Party Support: A Method for Long-Range Daily Election Forecasting from Opinion Polls.” *Journal of Elections Public Opinion and Parties* 25 (2): 1-22.

Garzia, D. 2014. *Personalization of Politics and Electoral Change*. New York: Palgrave

Giuliani, M. and S. A. Massari. 2017. “The economic vote at the party level: Electoral behaviour during the Great Recession.” *Party Politics* DOI: 10.1177/1354068817728214

Giuliani, M. and S.A. Massari. 2018. *It’s the Economy, Stupid. Votare in tempo di crisi*. Bologna: il Mulino.

Gramlich, J. 2017. “Q&A: Pew Research Center’s president on key issues in U.S. polling”. Pew Research Center. *Fact Tank. News in Numbers*: June, 16.

Hix, S., and M. Marsh. 2011. “Second-order effects plus pan-European political swings: An analysis of European Parliament elections across time.” *Electoral Studies* 30 (1): 4-15.

Hobolt, S. B., and J. Wittrock. 2011. “The Second-order Election Model Revisited: An Experimental Test of Vote Choices in European Parliament Elections.” *Electoral Studies* 30 (1): 29-40.

Jackman, S. 2005. “Pooling the Polls Over an Election Campaign.” *Australian Journal of Political Science* 40 (4): 499–517.

Jennings, W., and C. Wlezien. 2016. “The Timeline of Elections: A Comparative Perspective.” *American Journal of Political Science* 60 (1): 219–233.

Jennings, W., and C. Wlezien. 2018. “Election polling errors across time and space.” *Nature Human Behaviour* 2 (4): 276–283.

Kam, C. D., and R. J. Franzese. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor: The University of Michigan Press.

Leiter, D, A. Murr, E. Rascón Ramírez and M. Stegmaier. 2018. “Social networks and citizen election forecasting: The more friends the better.” *International Journal of Forecasting* 34 (2): 235–248.

Lewis-Beck, M. S. 2005. “Election Forecasting: Principles and Practice.” *The British Journal of Politics and International Relations* 7 (2): 145-164.

Lewis-Beck, M. S. and C. Tien. 2000. “The Future in Forecasting: Prospective presidential models.” In *Before the Vote: Forecasting American National Elections*, edited by J. Campbell and J. Garand, 133-144. Thousand Oaks: Sage.

Lewis-Beck, M. S. and C. Tien. 2012. “Election forecasting for turbulent times.” *PS: Political Science and Politics* 45 (4): 625-629.

Lewis-Beck, M. S. and C. Tien. 2016. “The Political Economy Model: 2016 US Election Forecasts.” *PS: Political Science and Politics* 49 (4): 661-663.

Lewis-Beck, M.S. and R. Dassonneville. 2015. “Forecasting elections in Europe: synthetic models.” *Research & Politics* 2 (1): 1-11.

Lewis-Beck, M.S. and M. Stegmaier. 2007. “Economic models of voting.” In *The Oxford handbook of political behaviour*, edited by R.J. Dalton and H.D. Klingemann, 518-537. Oxford: Oxford U.P.

Lewis-Beck, M.S. and M. Stegmaier. 2014. “Election forecasting, scientific approaches.” In *Encyclopedia of social network analysis and mining*, edited by J. Rokne and R. Alhajj, 462-469, Heidelberg: Springer.

Litton K. 2012. “Party novelty and economic voting: the evidence from the EU parliamentary elections.” *Paper MPSA Conference*, 12–15 April, Chicago (Il). https://astro.temple.edu/~klitton/docs/PartyNoveltyAndEconomicVoting.pdf

Litton K. 2015. “Party novelty: conceptualization and measurement of party change.” *Party Politics* 21 (5): 712–725.

Mahoney, J., and G. Goertz. 2006. “A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research.” *Political Analysis* 14 (3): 227-249.

Marsh, M. 1998. “Testing the Second-Order Election Model after Four European Elections.” *British Journal of Political Science* 28 (4): 591–607.

Martin, E., M. Traugott, and C. Kennedy. 2005. “A Review and Proposal for a New Measure of Poll Accuracy.” *The Public Opinion Quarterly* 69 (3): 342-369.

Noelle-Neumann, E. 1993. *The Spiral of Silence: Public Opinion - Our Social Skin*. Chicago: University of Chicago Press.

Reif, K. 1984. “National Electoral Cycles and European Elections, 1979 and 1984.” *Electoral Studies* 3 (3): 244-255.

Reif, K., and H. Schmitt. 1980. “Nine Second-Order National Elections. A Conceptual Framework for the Analysis of European Election Results.” *European Journal of Political Research* 8 (1): 3-44.

Schmitt, H., and E. Teperoglou. 2015. “The 2014 European Parliament Elections in Southern Europe: Second-Order or Critical Elections?.” *South European Society and Politics* 20 (3): 287-309.

Schmitt, H., and I. Toygür. 2016. “European Parliament Elections of May 2014: Driven by National Politics or EU Policy Making?.” *Politics and Governance* 4 (1): 167-181.

Stegmaier, M., M.S. Lewis-Beck, and B. Park. 2017. “The VP-function: A review.” In *The SAGE Handbook of Electoral Behaviour*, edited by K. Arzheimer, J. Evans and M.S. Lewis-Beck, 584-605, London: Sage.

van der Eijk, C., and H. Schmitt. 2009. “Legitimacy and Electoral Abstentions in European Parliament Elections.” In *The Legitimacy of the European Union after enlargement*, edited by J. Thomassen, 208-224. Oxford: Oxford University Press.

Voss, S. D., A. Gelman, and G. King. 1995. “The Polls. A Review. Preelection Survey Methodology: Details from Eight Polling Organizations, 1988 and 1992.” *Public Opinion Quarterly* 59 (1): 98-132.

Waldrop, M.M. 1993. *Complexity: the emerging science at the edge of order and chaos.* New York: Touchstone

Weber, T. 2007. “Campaign Effects and Second-Order Cycles: A Top-Down Approach to European Parliament Elections.” *European Union Politics* 8 (4): 509–36.

Weber, T. 2011. “Exit, Voice, and Cyclicality: A Micrologic of Midterm Effects in European Parliament Elections.” *American Journal of Political Science* 55 (4): 906–921

[1] In truth, some of these results have been attributed to the inability sufficiently to detail the geographical distribution of vote intentions, so that the disproportionality of the electoral system artificially inflated the gap between overall prediction and actual political result. This was especially the case of the discrepancy between popular vote and Electoral College in the US presidential election. A very recent article by Jennings and Wlezien (2018), surveying polling errors in 45 countries for 75 years, demonstrate that there is not a historical downward trend in the accuracy of predictions, thus suggesting a diffused and media-inflated misperception of recent misses. The results of their analysis, checking for the origin of absolute polling errors, does not invalidate our own directional and election-specific results.

[2] In the article we have consistently followed the distinction between (pollsters') predictions and (models') forecasts. Yet, if "forecasting requires more than curve-fitting" (Lewis-Beck and Tien 2000, 98), since it is a theory-driven process (Lewis-Beck 2005), this does not mean that polls simply require the computation of frequencies. The way in which the sample is constructed and balanced, non-respondents are treated, and, eventually, estimates are produced, require a set of conscious choices (Fisher et al. 2011) that can certainly profit from a theory-laden interpretation of the process, and a consolidated knowledge of the context-specific features of that election.

[3] The list of polls, mostly run as CATI and typically with four times more contacts than actual respondents, was completed by drawing on three different sources: a dataset published shortly after the election by @electionista, the Pollwatch 2014 website run by Votewatch Europe http://www.votewatch.eu/ , and the 28 pages of Wikipedia in their original language devoted to the electoral appointment. We complemented that information by using several sources, ranging among direct polls reports, newspapers or weekly journals, EU official results http://www.europarl.europa.eu/elections2014-results/en/election-results-2014.html, national election data and indices from the Parlgov project (Döring and Manow 2015). The surveys asked for vote intentions in the imminent European election, and thus respondents' answers should have had a built in second-order effect, as we could verify for some polls that asked in parallel the preferences for the European and a hypothetical national ballot. The dataset is available in the author's personal webpage.

[4] There are just five occurrences for which, at the level of precision of our data, pollsters perfectly predicted the result of a party. Given the negligible amount of these cases, the incidence of a negative error can be simply considered as the mirror image of positive errors, without needing to test them separately.

[5] We acknowledge that this is a rather drastic simplification of the complex issue of what really constitutes a "new" party, how to treat cases of split/merger of preceding groups (Bolleyer and Bytzek 2013; Emanuele and Chiaramonte 2016), and if novelty is a dichotomous property or is better captured by a continuous measure (Barnea and Rahat 2011; Litton 2012; 2015). We cannot further develop our conceptual analysis in this direction, but in the empirical part we will check the robustness of alternative operationalizations in order to shed some light on the actual meaning of "new" for the present hypotheses.

[6] Polls often estimate the support of only a subset of cases, sometimes reporting collectively as "others" the votes for the remaining parties, not considered in our analysis. This justifies the positive error for both government and opposition parties.

[7] For the analysis of bias, we also experimented with a single multinomial logistic model. Yet our hypothesis does not actually translate into contrasting the opposite type of biases against a null baseline, but each type against the remaining options together (i.e. no bias plus wrong ones). For this reason, we preferred separate logistic models. Yet, the results are mostly similar, with the partial exception of Euroscepticism, which seems to depress both type of biases against the null hypothesis.

[8] As anticipated in note 5, we also operationalized novelty in alternative ways instead of its simply being absent in the previous, and usually nearest, general election. We checked for parties taking part in their first European ballot, or for having just one previous national electoral experience. We also tried a four-point scale, from zero to three, in which "newness" was the reverse of a count variable measuring the number of previous national and European ballots in which the party took part. The results of these alternative operationalizations are reported in the online appendix (Table A.8). Interestingly, none of these substitutes proves to have a significant coefficient. It is not being new just for the European election that counts for the errors, and, for that matter, the "virginity" seems to be immediately lost after the first experience both in the ballot and in the polls. This indirectly confirms the importance of a contextual knowledge of the specific election, and of the actors that take part in it.

[9] In fact, the same theory is a usual component of any forecasting effort during and after the Great Recession, including the mentioned US presidential election (Lewis-Beck and Tien 2012; 2016; Lewis-Beck and Stegmaier 2014; Lewis-Beck and Dassonneville 2015).