Abstract

The present research note contributes to the (citizen) forecasting literature by leveraging vote expectation data from close to 3,200 district-level races in four countries (i.e., Canada, France, Germany, and Great Britain) in order to assess the relative merits of two competing views of the 'wisdom of crowds' hypothesis: the democratic view and the technocratic view. More precisely, the paper addresses the following question: Can we improve citizens' forecasts of election outcomes by weighting voters' expectations according to their level of expertise?

Keywords: aggregation, citizen forecasting, competence, political knowledge, voter expectations, wisdom of crowds

A Technocratic View of Election Forecasting: Weighting Citizens' Forecasts According to Competence

Philippe Mongrain

International Journal of Public Opinion Research, accepted

2020

Introduction

According to the 'wisdom of crowds' (WOC) principle, "under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them. [...] Even if most of the people within a group are not especially well-informed or rational, it can still reach a collectively wise decision" (Surowiecki, 2004, pp. xiii–xiv). The benefits of collective judgment have already been demonstrated multiple times in the fields of psychology, statistics, and management science (see, for example, Clemen 1989; Davis-Stober, Budescu, Dana, and Broomell 2014; (Dunning, 2007, pp. 84–86); (Stewart, 2001, pp. 95–96); Yaniv 2004). In political science more specifically, the WOC principle has been invoked as the main explanation underlying the accuracy of voters' collective expectations about electoral outcomes. A number of studies in the United States, the United Kingdom, and Canada have shown that most citizens were able to correctly predict which party or candidate would win in local, regional, or national elections and that the aggregation of individual estimates increased the likelihood of a correct forecast (see, for example, Graefe 2014; Lewis-Beck and Stegmaier 2011; Miller, Wang, Kulkarni, Poor, and Osherson 2012; Murr 2011; 2015; 2016; Temporão, Dufresne, Savoie, and van der Linden 2019). In other words, the proportion of correctly predicted outcomes was always higher than the proportion of correct individual forecasts.

Murr (2015) has also demonstrated that delegating the forecasting task to individuals with 'competence-enhancing' characteristics and weighting their forecasts according to their level of expertise increased the proportion of correctly predicted state outcomes in American presidential elections. Murr's (2015) findings question the 'democratic view' of the WOC principle—i.e., that every member of the group should be treated with equal importance—and point toward the potentially superior efficiency of a 'technocratic view,' in which the judgments of sophisticated or competent individuals are attributed greater weight. However, the limited amount of work devoted to studying whether adjusting forecasts based on competence has an impact on accuracy offers only mixed evidence (see Babad 1995; Ganser and Riordan 2015; Murr 2015).

The present research note contributes to the (citizen) forecasting literature by leveraging vote expectation data from more than 3,200 district-level races in four countries (i.e., Canada, France, Germany, and Great Britain) in order to assess the relative merits of two competing views of the WOC hypothesis: the democratic view and the technocratic view. More precisely, the paper addresses the following question: Can we improve citizens' forecasts of election outcomes by weighting voters' expectations according to their level of expertise?

The Wisdom of Crowds and Aggregation

The WOC principle is mostly derived from Condorcet's (1785) jury theorem, which states that, under certain conditions, the probability of a group coming to a correct decision tends toward unity as the group increases in size. This principle is closely related to what others have called the 'miracle of aggregation.' The main argument is that statistical aggregation cancels out errors in individual judgments (Page & Shapiro, 1992). When individuals are asked to make a forecast of some sort, simply aggregating all forecasts is obviously not the only approach in order to obtain the best possible prediction. Identifying the most competent member of the crowd based on some predefined criteria—or, put another way, finding 'the best and the brightest'—and relying solely on the estimate of that member is another option (Budescu & Chen, 2015). However, according to Davis-Stober et al. (2014), it is rarely justified to ignore the aggregated judgment of a group in favour of the single opinion formulated by an expert. As stated by Budescu and Chen (2015; see also (Hertwig, Pleskac, & Pachur, 2019, p. 254)), a possible compromise between the WOC approach and the 'lone expert' approach involves retaining only the best subset of judges in a group (i.e., a 'select crowd').

The evidence regarding the usefulness or efficiency of the select-crowd approach for election forecasting is somewhat mixed. In the case of American presidential elections, Murr (2015) has shown that discriminating between respondents based on past performance makes for more accurate forecasts. According to Murr (2015), the democratic principle of 'one person, one vote,' although necessary when aggregating preferences, might not be ideal when it comes to expectations. In fact, citizens' forecasts "can become even better when the task of forecasting is delegated to the most competent citizens, and their forecasts are weighted depending on their competence" (Murr, 2015, p. 917). In the context of the 2013 German federal election, however, Ganser and Riordan (2015) have found that 'high-ability subgroups' (i.e., the well informed, the academically educated, and the highly knowledgeable) were not better at forecasting the vote shares of political parties than the whole sample of citizen forecasters. After surveying a sample of Israeli voters before the 1992 Knesset election, Babad (1995, p. 294) came to the conclusion that "existing knowledge had a very small effect on respondents' realistic estimation [of election results] and a negligible effect on wishful thinking, whereas the experimental provision of information had no effect at all." In fact, there are indications in the literature that increasing group diversity might be more profitable than seeking out expertise (see Davis-Stober et al. 2014; Ganser and Riordan 2015; Hong and Page 2004; Larrick, Mannes, and Soll 2012).

To verify the claims of the technocratic view, I first identify potential indicators of expertise, and I then proceed to weigh voters' expectations based on these indicators. The following section covers the data and methodological framework employed to achieve these tasks.

Data and Methods

Dataset

The necessary data were drawn from 12 election studies, that is the British Election Study (BES) Internet Panel, the 2000, 2004, 2006 and 2008 Canadian Election Studies (CES), and seven electoral surveys from the Making Electoral Democracy Work (MEDW) project. These studies provide the needed information on voters and voters' electoral expectations at the district level for 12 different elections (i.e., five national and two regional elections in Canada, one national legislative election in France, one national election and one regional election in Germany as well as two national elections in Great Britain). Furthermore, the BES Internet Panel and the 2004–2006 panel of the CES prove valuable for the purpose of the present research note: the BES and CES are among the very rare studies to have questioned the same individuals about their electoral expectations at the district level for more than a single election which both allows me (1) to assess the predictive power of past performance on subsequent forecasts and (2) to use past performance as a weighting mechanism.

Methods

The main objective of this research note is to evaluate the relative merits of weighting forecasts according to potential indicators of competence. To fulfill this objective, I proceed in two related steps. First, the association between potential predictors of forecasting ability and forecasting ability itself is assessed through random intercept multilevel linear and logistic regression analyses (depending on the nature of the dependent variable) controlling for political preferences as well as sociodemographic and contextual variables. Multilevel analysis was chosen since voters are nested within districts. Second, the most promising variables (which turned out to be political knowledge and past performance) were used to weigh respondents' forecasts.

Variables[1]

Forecasting ability (DV)

In the 2006 and 2008 CES, respondents were asked to mention from which party they believed the winning candidate would be. Hence, the accuracy of voter expectations was measured using a binary indicator coded 1 for correct forecasts and 0 otherwise. MEDW respondents were asked to evaluate the winning probability of each party candidate in their own district on a 0–10 scale, where 0 meant 'no chance at all' and 10 meant 'certain to win.' A similar question was asked in the BES Internet Panel as well as the 2000 and 2004 CES, this time on a 0–100 scale. The accuracy of voter expectations was measured by using a summation of (half) Brier scores. The Brier score is the sum of the squared differences between the predicted probability (i.e., between 0 and 1) of an event occurring and the actual outcome coded 1 if the events occurs and 0 if it does not (see (Jackson, 2018, p. 623)). Since respondents were asked to make a probability forecast for multiple candidates in their district, their forecasting accuracy can be measured by adding the Brier score of each potential outcome, a measure I later refer to as the 'cumulative Brier score' (CBS) index. Equation 1 below shows how this index (which ranges from -1 to 1) is computed.[2]

$$CBS_{ij} = 1 - \sum_{m=1}^{k} (f_{mj} - o_{mj})^2 \tag{1}$$

in which CBS_{ij} is the cumulative Brier score for respondent i in district j, f_{mj} is the forecasted probability of winning for party m in district j, o_{mj} is the actual outcome of the election for party m in district j, and k is the number of parties for which a probability forecast is made.

Predictors (IVs)

Past research (Dolan & Holbrook, 2001; Lewis-Beck & Tien, 1999; Meffert, Huber, Gschwend, & Pappi, 2011; Mongrain, 2021; Murr, 2011) suggests that variables related to the level of information possessed by individuals such as interest for the campaign, news attentiveness, and factual knowledge about politics can have a positive impact on citizens' forecasting accuracy. An indicator for each of these variables was included in the models.[3] Political knowledge questions in the 2000, 2004, 2006, and 2008 CES were open-ended. In the MEDW surveys, the knowledge items involved an association task (i.e., matching leaders and campaign promises with the appropriate parties) akin to a multiple choice format. All knowledge items either probed respondents' familiarity with political figures or their awareness of parties' election pledges. Each answer was coded as either correct (1) or incorrect (0). Political knowledge is often measured using an additive scale (see, for example, Daoust, Durand, and Blais 2020; Delli Carpini and Keeter 1991; 1993). This approach, however, has a number of drawbacks, the most important being the lack of distinction between items: one can reasonably expect that a set of factual questions will include items of varying difficulty. However, additive scales assume that every item have similar properties. Item response theory (IRT) appears as a valuable alternative to summated test scores (DeMars, 2010). IRT was used to compute an ability (θ) score for each respondent (i.e., their level on the latent trait of political knowledge). Note that due to space limitations, methodological details related to the computation of ability scores are provided in the appendix (section B). Finally, past performance, which is available only for the BES Internet Panel respondents and panel respondents of the 2006 CES, was measured using these respondents' CBS values from the previous election. Conventional sociodemographic characteristics (i.e., gender, age, and education), respondents' partisan leanings (a five-point scale ranging from 'strong loser PID' to 'strong winner PID'), time of interview, competitiveness (measured using the effective number of electoral parties or ENEP), and incumbent reelection (when there was no redistricting between elections) were added to the models as controls. An interaction term between partisan leaning and knowledge was also added to the models to account for the possibility that more knowledgeable respondents are less susceptible to letting their preferences orient their expectations (Meffert et al., 2011).

Aggregation and weighting

Following Murr (2011), I rely on two different aggregation procedures—i.e., plurality voting (PV) and range voting (RV). In the case of plurality voting, the party with the most forecasts in its favour is predicted to win the district. Forecasts are coded 1 if a plurality of voters inside the district predicted the correct outcome and 0 otherwise. Although PV is straightforward, it has the disadvantage of discarding a lot of information. A respondent whose highest expectation is equal to 40% is treated in the same manner as a respondent whose highest expectation is equal to 90%. However, it is safe to say that one of the two is more confident about his or her forecast than the other. The range voting procedure takes into account respondents' level of confidence (Murr, 2011, p. 774). The RV procedure consists in computing the sum of expectation scores for each party across voters in a district. Consequently, district forecasts are coded 1 if the winning party has the *highest sum of expectations* and 0 otherwise.[4] Section E of the appendix provides an example that shows how both aggregation mechanisms (i.e., PV and RV) work. For every election (with the exception of British elections), the ability (political knowledge) score was used as a weighting device in two different ways. First, the percentage of highest expectations for a party in a district was weighted by the mean (normalized) ability score of respondents (within the same district) who gave their highest expectation to that party. The party with the highest weighted percentage was considered as the forecasted winner. I later refer to this weighting procedure as 'weighted plurality voting' or WPV.[5] For example, take a district where three parties compete (i.e., Party A, Party B, and Party C). Party A won the election in that district. Imagine 40% of respondents in the district believed Party A would win, 50% believed Party B would win, and 10% believed Party C would win. According to the plurality rule, Party B is the forecasted winner. The mean ability score of respondents who forecasted the victory of Party A is 0.7, it is 0.5 for those who expected Party B to win, and 0.4 for those who thought Party C would prevail. The weighted percentages (once converted back to sum 100%) are 49.1% for Party A, 43.9% for Party B, and 7.0% for Party C. Hence, once the weighting procedure is applied, Party A (rather than Party B) is the forecasted winner. Equation 2 shows how these weighted percentages are obtained (the weighting component is in bold):

$$wp_{mj} = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left(\frac{\boldsymbol{\theta_{ij}-\theta_{min}}}{\boldsymbol{\theta_{max}-\theta_{min}}} \right)}{\boldsymbol{n}} \times p_{mj} \tag{2}$$

> in which wp_{mj} is the weighted percentage of respondents who believe party m in district j will win, θ_{ij} is the ability score of respondent i in district j, θ_{min} is the minimum value on the ability scale, θ_{max} is the maximum value on the ability scale, p_{mj} is the percentage of respondents who believe party m in district j will win, and n is the number of respondents who forecasted party m's victory in the district.

The second weighting scheme is similar to the range voting procedure, the only difference being that individual probability estimates of candidates' chances are replaced by the ability scores of respondents. In other words, the party with the highest sum of (normalized) ability scores among the respondents who forecasted its victory is considered as the forecasted winner. I later refer to this weighting procedure as 'plurality-range voting' or PRV. For the BES and the 2004–2006 CES panel respondents, the same two weighting schemes (i.e., WPV and PRV) were also applied by replacing the ability scores of respondents with their (normalized) previous election CBS value as a measure of past performance.

Results

Figure 1 shows regression coefficients for each election in which forecasting ability was measured using the CBS index. Figure 2 shows average marginal effects (AME) for the 2006 and 2008 Canadian federal elections (as well as the 2004–2006 CES panel respondents) in which forecasting ability was measured using a binary (correct/incorrect) indicator.[6] Interest for the election is statistically significant in only three elections while news attentiveness is never significant. This is perhaps not surprising since these indicators might be more relevant for the national- or regional-level contest than they are for the district race. In line with past research, the results also suggest the presence of wishful thinking (citizens who identify with the eventual winner are more likely to predict its victory than those who side with one of the losing parties). As can also be seen, elections are generally harder to predict in more competitive districts and easier to predict where the incumbent party candidate is reelected. Political knowledge is statistically significant in seven of the 10 elections for which political knowledge items were available. The influence of political knowledge is, however, quite small. Results from the 2017 and 2019 British general elections as well as the 2004–2006 CES panel show that past performance (as measured through the respondents' CBS value from the previous election) is positively associated with forecasting ability. Once again, however, the association is weak.

[Figure 1 about here]

[Figure 2 about here]

Figure 3 shows the percentage of districts correctly predicted once individual forecasts are aggregated (using the PV and RV methods) and weighted (either through WPV or PRV) to take into account respondents' political knowledge or past performance. Although it seems difficult to make a general conclusion that equally applies to all elections, we can see that the simplest and most common way of aggregating expectations (i.e., PV) almost always has the lowest percentage of correctly predicted district-level outcomes. With the exception of the 2013 German federal election, where WPV leads to an important decrease in accuracy, weighting produces small to sizable increases in the percentage of correct forecasts depending on the election. Interestingly, RV almost consistently performs better than WPV and cannot be dismissed as inferior to PRV. The confidence intervals show, however, that the difference between methods in the proportions of correct forecasts is almost never significant. I would argue, however, that this is mostly due to the relatively small N for most elections. Dismissing the results as meaningless based on the significance of two-sample tests of proportions would be somewhat premature: the results suggest that one would almost always do better by using RV instead of PV if expectations were measured in a probabilistic fashion and that, absent probability estimates, weighting according to competence would almost always prove preferable than PV. The average increase is 8.21 percentage points between PV and RV, 10.95 percentage points between PV and PRV, and 3.03 percentage points between PV and WPV.

In many cases, these increases are far from trivial: for example, in the 2019 British general election, the PRV procedure makes a difference of 76 additional correctly predicted outcomes compared to RV (i.e., 436 vs 360 districts). Furthermore, confidence intervals for pooled data show that the difference between PV and PRV is, overall, statistically significant (see section G of the appendix).

[Figure 3 about here]

Note that since all the elections considered here were contested in single-member districts according to the 'first-past-the-post' principle (with the exception of France[7]), we should not expect institutional differences to considerably influence the efficacy of the aggregation and weighting methods previously discussed. Furthermore, the limited number of cases per country hardly allows the detection of specific system-level patterns. It is true, however, that under different electoral laws, the proposed weighting schemes might not work as well.

Discussion and Conclusion

Although the literature on citizen forecasting is growing, few studies have considered the potential benefits of using voters' differential levels of competence or sophistication as a means of increasing the accuracy of electoral predictions. This research note compared two competing views of the WOC principle, that is the democratic view—which treats all forecasts equally—and the technocratic view—which considers that opinions should be weighted according to potential indicators of expertise. This paper has shown two things: (1) RV appears as a superior alternative to PV when expectations are measured on a probabilistic scale and (2) when expectations are recorded in a non-probabilistic fashion, as is often the case (e.g., by simply asking who will win), applying weights appear as a relatively simple way to improve upon conventional (unweighted) aggregation. Both political knowledge and past performance seem to be potent markers of expertise. Since most elections under study had a relatively small number of observations per district, I was unable to assess the benefits of *delegating* the forecasting task to the most competent individuals—an option that, combined with weighting, could prove quite successful. Finally, the study of electoral expectations would greatly benefit from the designing of surveys including a panel component (to efficiently track expertise) and a large number of respondents in *every* district.

References

Babad, E. (1995). Can accurate knowledge reduce wishful thinking in voters' predictions of election outcomes? *Journal of Psychology*, *129*(3), 285–300. doi: 10.1080/00223980.1995.9914966

Budescu, D. V., & Chen, E. (2015). Identifying expertise to extract the wisdom of crowds. *Management Science*, *61*(2), 267–280. doi: 10.1287/mnsc.2014.1909

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583. doi: 10.1016/0169-2070(89)90012-5

Condorcet, m. d., Marie-Jean-Antoine-Nicolas de Caritat. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: Imprimerie Royale.

Daoust, J.-F., Durand, C., & Blais, A. (2020). Are pre-election polls more helpful than harmful? Evidence from the Canadian case. *Canadian Public Policy / Analyse de politiques*, *46*(1), 175–186. doi: 10.3138/cpp.2019-011

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision*, *1*(2), 79–101. doi: 10.1037/dec0000004

Delli Carpini, M. X., & Keeter, S. (1991). Stability and change in the U.S. public's knowledge of politics. *Public Opinion Quarterly*, *55*(4), 583–612. doi: 10.1086/269283

Delli Carpini, M. X., & Keeter, S. (1993). Measuring political knowledge: Putting first things first. *American Journal of Political Science*, *37*(4), 1179–1206. doi: 10.2307/2111549

DeMars, C. (2010). *Item response theory*. Oxford: Oxford University Press.

Dolan, K. A., & Holbrook, T. M. (2001). Knowing versus caring: The role of affect and cognition in political perceptions. *Political Psychology*, *22*(1), 27–44. doi: 10.1111/0162-895X.00224

Dunning, D. (2007). Prediction: The inside view. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 69–90). New York: Guilford Press.

Ganser, C., & Riordan, P. (2015). Vote expectations at the next level: Trying to predict vote shares in the 2013 German federal election by polling expectations. *Electoral Studies*, *40*(Dec.), 115–126. doi: 10.1016/j.electstud.2015.08.001

Graefe, A. (2014). Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly*, *78*(S1), 204–232. doi: 10.1093/poq/nfu008

Hertwig, R., Pleskac, T. J., & Pachur, T. (2019). *Taming uncertainty*. Cambridge, MA: MIT Press.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385–16389. doi: 10.1073/pnas.0403723101

Jackson, N. (2018). The rise of poll aggregation and election forecasting. In L. R. Atkeson & R. M. Alvarez (Eds.), *The Oxford handbook of American elections and political behavior* (pp. 609–632). Oxford: Oxford University Press.

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 227–242). New York: Psychology Press.

Lewis-Beck, M. S., & Stegmaier, M. (2011). Citizen forecasting: Can UK voters see the future? *Electoral Studies*, *30*(2), 264–268. doi: 10.1016/j.electstud.2010.09.012

Lewis-Beck, M. S., & Tien, C. (1999). Voters as forecasters: A micromodel of election prediction. *International Journal of Forecasting*, *15*(2), 175–184. doi: 10.1016/S0169-2070(98)00063-6

Meffert, M. F., Huber, S., Gschwend, T., & Pappi, F. U. (2011). More than wishful thinking: Causes and consequences of voters' electoral expectations about parties and coalitions. *Electoral Studies*, *30*(4), 804–815. doi: 10.1016/j.electstud.2011.08.001

Miller, M. K., Wang, G., Kulkarni, S. R., Poor, H. V., & Osherson, D. N. (2012). Citizen forecasts of the 2008 U.S. presidential election. *Politics and Policy*, *40*(6), 1019–1052. doi: 10.1111/j.1747-1346.2012.00394.x

Mongrain, P. (2021). Did you see it coming? Explaining the accuracy of voter expectations for district and (sub)national election outcomes in multi-party systems. *Electoral Studies*, *71*.

doi: 10.1016/j.electstud.2021.102317

Murr, A. E. (2011). 'Wisdom of crowds'? A decentralised election forecasting model that uses citizens' local expectations. *Electoral Studies*, *30*(4), 771–783. doi: 10.1016/j.electstud.2011.07.005

Murr, A. E. (2015). The wisdom of crowds: Applying Condorcet's jury theorem to forecasting US presidential elections. *International Journal of Forecasting*, *31*(3), 916–929. doi: 10.1016/j.ijforecast.2014.12.002

Murr, A. E. (2016). The wisdom of crowds: What do citizens forecast for the 2015 British general election? *Electoral Studies*, *41*(Mar.), 283–288. doi: 10.1016/j.electstud.2015.11.018

Page, B. I., & Shapiro, R. Y. (1992). *The rational public: Fifty years of trends in Americans' policy preferences*. Chicago: University of Chicago Press.

Stewart, T. R. (2001). Improving reliability of judgmental forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 81–106). Norwell, MA: Kluwer Academic Publishers.

Surowiecki, J. (2004). *The wisdom of crowds*. London: Abacus.

Temporão, M., Dufresne, Y., Savoie, J., & van der Linden, C. (2019). Crowdsourcing the vote: New horizons in citizen forecasting. *International Journal of Forecasting*, *35*(1), 1–10. doi: 10.1016/j.ijforecast.2018.07.011

Yaniv, I. (2004). The benefit of additional opinions. *Current Directions in Psychological Science*, *13*(2), 75–78. doi: 10.1111/j.0963-7214.2004.00278.x

Footnotes

[1]Data sources, coding information, and descriptive statistics can be found in the appendix (sections A, C, and D respectively).

[2]The sum of Brier scores is subtracted from 1 so positive values correspond to better forecasts.

[3]However, the items measuring political knowledge in the BES Internet Panel were very different in nature from the items found in the other surveys (see footnote 4 in the appendix). Hence, political knowledge was not included in British models.

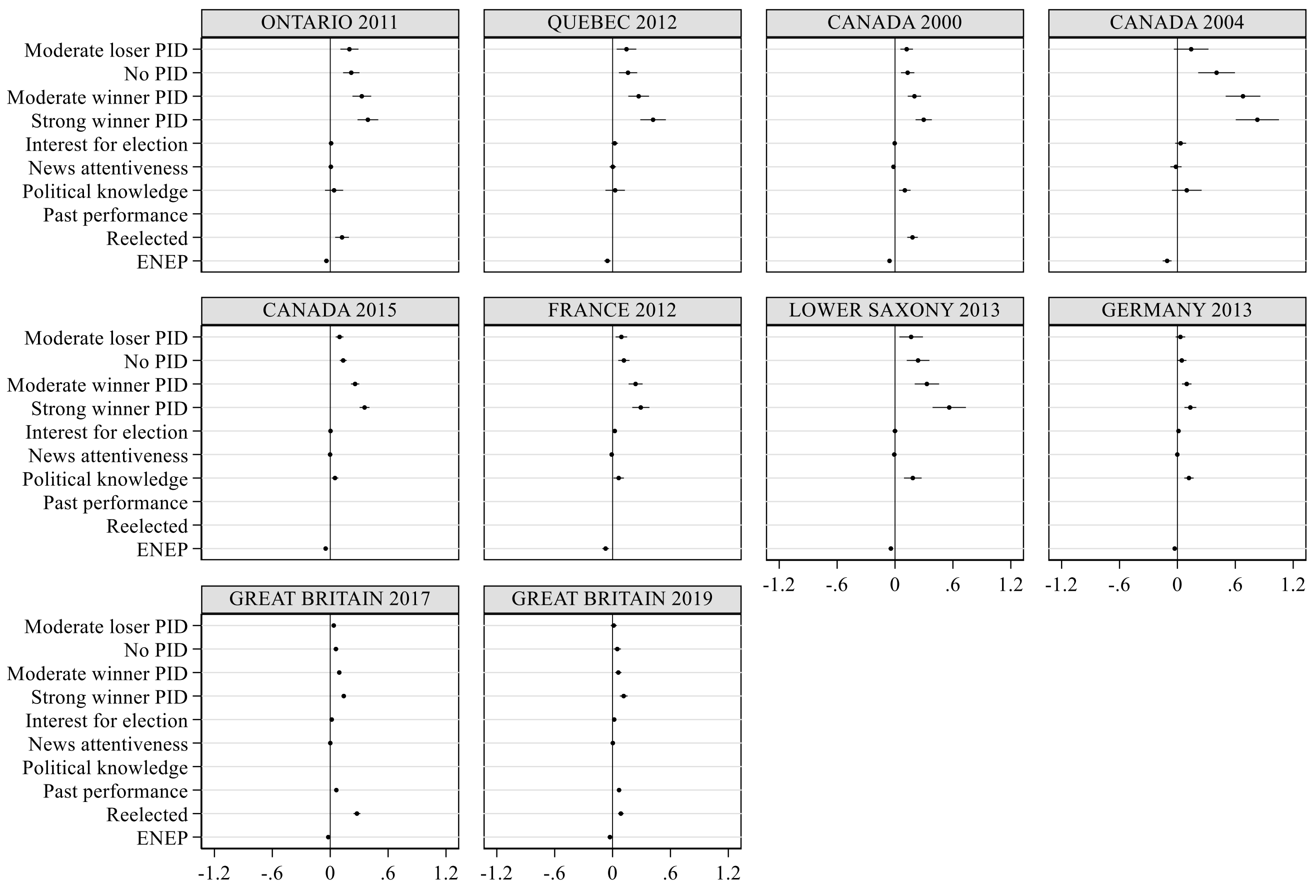
[4]The range procedure was not applied to the 2006 and 2008 CES since it requires probability estimates.

[5]Note that the weighting procedure *cannot* be applied to the range voting rule since the sum of expectations for each party involves the answers of every respondent in the district (hence, the weight would be exactly the same for each sum of expectations).

[6]See section F of the appendix for complete regression results. Section H contains information related to the replication of results.

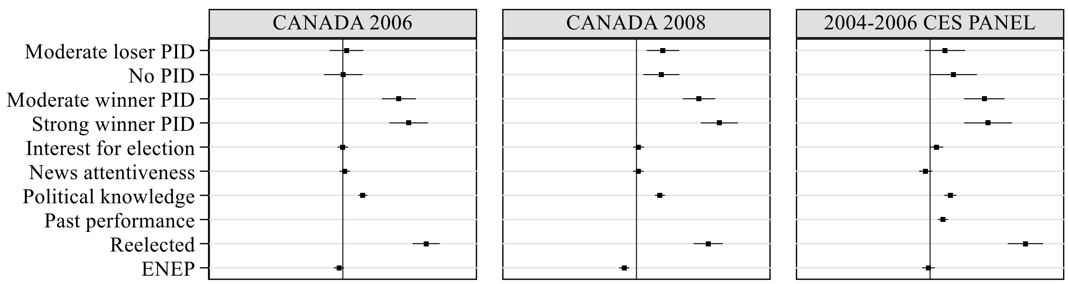
[7]In France, elections take place under a two-round system. MEDW French respondents were asked to make a forecast for the second round of the election *before* the first round. This might explain the somewhat lower proportions of correct forecasts for this election.

FIGURE 1. Linear prediction (fixed portion)



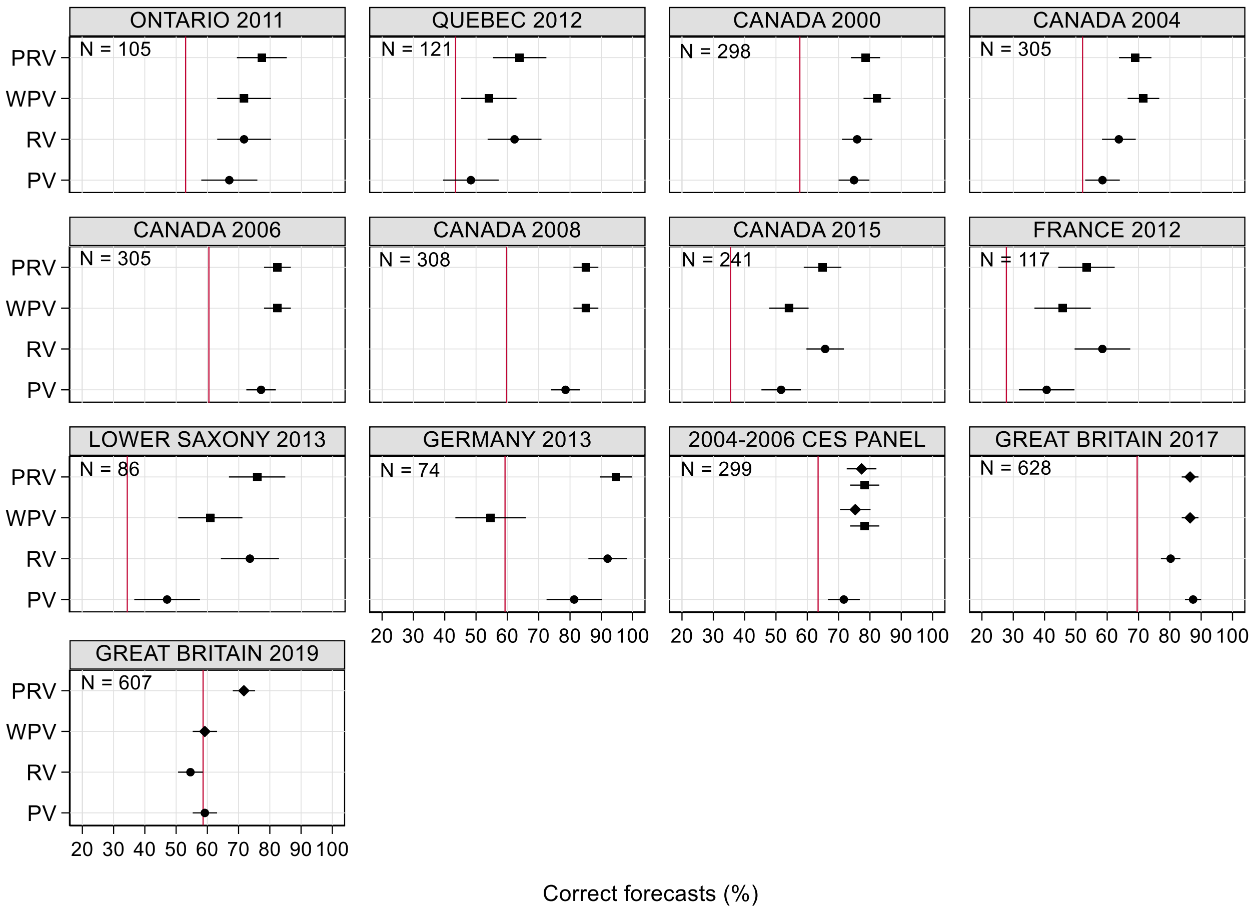
Note. The horizontal lines represent the 95% confidence intervals. The reference category for PID is 'strong loser PID.' Continuous independent variables with no natural metric (i.e., all variables with the exception of reelection, gender, and age) were standardized (i.e., mean = 0, sd = 1) to ease comparisons. Interaction coefficients as well as sociodemographic and time controls are not shown.

FIGURE 2. Average marginal effects



Note. The horizontal lines represent the 95% confidence intervals. The reference category for PID is 'strong loser PID.' Continuous independent variables with no natural metric (i.e., all variables with the exception of reelection, gender, and age) were standardized (i.e., mean = 0, sd = 1) to ease comparisons. Sociodemographic and time controls are not shown.

FIGURE 3. Aggregated and weighted forecasts



Note. The vertical red line shows the percentage of correct individual-level forecasts for reference. N = number of districts. Circles = no weight; Squares = political knowledge weight; Diamonds = past performance weight.