International Journal of Forecasting 31 (2015) 916–929

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The wisdom of crowds: Applying Condorcet's jury theorem to forecasting US presidential elections

Andreas E. Murr

Department of Politics and International Relations, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, UK

ARTICLE INFO

Keywords:
Citizen forecasting
Combining forecasts
Condorcet's jury theorem
Election forecasting
Election surveys
Weighting

ABSTRACT

Increasingly, professional forecasters rely on citizen forecasts when predicting election results. Following this approach, forecasters predict the winning party to be the one which most citizens have said will win. This approach predicts winners and vote shares well, but related research has shown that some citizens forecast better than others. Extensions of Condorcet's jury theorem suggest that naïve citizen forecasting can be improved by delegating the forecasting to the most competent citizens and by weighting their forecasts by their level of competence. Indeed, doing so increases both the accuracy of vote share predictions and the number of states forecast correctly. Allocating the state's electoral votes to the candidate who the most weighted delegates say will win yields a simple but successful forecasting model of the US Presidency. The 'wisdom of crowds' model predicts eight presidential elections out of nine correctly. The results suggest that delegating and weighting provide easy ways to improve citizen forecasting.

1. Introduction

Democracies repeatedly ask citizens to elect the candidates, parties, or policies they want or consider to be best. Accordingly, considerable amounts of effort have been spent on forecasting the results of these elections. These attempts fall into two broad categories. Forecasting models in the first category focus on preferences. Forecasting models in this category usually either elicit vote intentions in surveys before the election and use the results to predict vote choice on election day (e.g., Linzer, 2013); or apply a general theory of voting behaviour — such as the idea that voters reward government for a good economy and punish it for a bad one — to forecast how voters will behave in a particular election (e.g., Abramowitz, 2012).

Forecasting models in the second category focus on judgements. They elicit vote expectations before the election and predict the winning candidate to be the one which most citizens forecast will win (e.g., Lewis-Beck & Skalaban, 1989). Because these models follow the democratic process by letting the simple majority of citizens decide which candidate is predicted to win, their predictive accuracy informs the debate about democracy itself.

The initial studies on citizen forecasting tested whether citizen forecasts predicted at all, and they did. They predicted well in the US when forecasting which candidate would win the Presidency (Lewis-Beck & Skalaban, 1989; Lewis-Beck & Tien, 1999) and which candidate would win a state (Graefe, 2014; Rothschild & Wolfers, 2013); they also predicted well in Great Britain when forecasting which party would win the Prime Ministership (Lewis-Beck & Stegmaier, 2011) and which party would win in each constituency (Murr, 2011).

More recent studies have even shown citizen forecasts to predict better than other approaches at the state and national levels. At the state level, Rothschild and Wolfers (2013) compared the relative accuracies of citizen forecasts and vote intentions. Based on data from the American National Election Studies (ANES) from 1952 to 2008,

E-mail address: andreas.murr@politics.ox.ac.uk.

http://dx.doi.org/10.1016/j.ijforecast.2014.12.002
0169-2070/

they found that citizen forecasts predicted more states correctly than did vote intentions in eight presidential elections out of ten. Similarly, Miller, Wang, Kulkarni, Poor, and Osherson (2012) compared the relative accuracies of citizen forecasts and the prediction market *Intrade*. Using a unique survey during the 2008 US presidential election, they showed that citizen forecasts predicted state outcomes more accurately than the prediction market.

At the national level, Graefe (2014) compared the relative accuracies of several approaches for predicting US presidential elections. The comparison pitted citizen forecasting against quantitative models, vote intentions, prediction markets, and election experts. Comparing their relative accuracies in the seven elections between 1988 and 2012, he found that citizen forecasts performed best in predicting election winners and vote shares.

Although the studies of citizen forecasting cited above aggregated judgements, they implicitly followed two aggregation principles for preferences in democratic elections: universal suffrage and 'one person, one vote'. The above-cited studies followed these principles by using all citizen forecasts and by giving them equal weights. These two principles are indispensable when aggregating preferences; however, it may be beneficial to dispense with them when aggregating judgements. Citizen forecasting 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.

Both delegating and weighting require a measure of forecasting competence, and the following study therefore proposes a new method of measuring it. The method identifies the forecasters that have forecast well in the past, estimates the probability of them being correct in the current election, and takes this estimate as a measure of their competence. It then delegates the forecasting task to the most competent members and weights their forecasts by their competence. Delegating and weighting increase the average probability of a correct forecast. Accordingly, the method improves citizen forecasts without having to interview more citizens. Instead, it improves forecasts based on the available data. The reason for the success of delegating and weighting lies in Condorcet's jury theorem and its extension, which highlight cases when groups outperform individuals in binary choice tasks (Condorcet, 1785, 1994).

2. Linking citizen forecasts to Condorcet's jury theorem

Murr (2011) was the first to identify the link between citizen forecasting and Condorcet's jury theorem. The theorem features prominently in discussions about the advantages and disadvantages of democracy, but it also applies to group decision making more generally. Condorcet derived the theorem for groups which need to choose between two alternatives by a simple majority vote, i.e., by choosing the alternative that receives the most votes from its members. He assumed that members vote independently of one another and that each member has the same chance of choosing the correct alternative.

The theorem implies that if each group member has a greater than 0.5 chance of voting for the correct alternative, then the probability of a correct simple majority vote approaches unity as the group size increases to infinity ('wisdom of crowds'). At the same time, the reverse also holds. If each member has a smaller than 0.5 chance of voting for the correct alternative, then the probability of a correct simple majority vote approaches zero as the group size increases to infinity ('folly of crowds'). However, the above-cited studies on citizen forecasting suggest that citizens have a greater than 0.5 chance of forecasting the election winner correctly.

The aggregation effect emerges even in small groups. Consider a group of citizens who each forecast correctly 60% of the time. With five citizens, the group forecasts correctly 68% of the time; with 25 citizens it does so 85% of the time. By contrast, with citizens who all forecast correctly 70% of the time, the groups of five and 25 citizens forecast correctly 84% and 98% of the time, respectively.[1]

3. Generalising Condorcet's jury theorem

Condorcet considered groups which face a binary choice situation with members who have the same competence levels and vote independently of one another. Subsequent research has generalised the theorem to situations where there are more than two alternatives, members with heterogeneous competence levels, and dependent votes. In the present context, the most important generalisations are those that allow for heterogeneous competence and dependent votes.

An important generalisation of Condorcet's jury theorem allows citizens to vary in their competence levels. Grofman (1978) demonstrated that Condorcet's jury theorem still holds if the competence of the group members is distributed symmetrically with a mean of greater than 0.5. This theoretical generalisation is important in practice because Lewis-Beck and Skalaban (1989) and Lewis-Beck and Tien (1999) have shown that citizens do indeed vary in their forecasting competence levels.

It is also likely that the forecasts of different citizens will be correlated with each other. With random sampling, it is unlikely that survey respondents have observed how the others forecasted, or that they have talked to each other to exchange information or to convince the other about who will win. Nevertheless, respondents might share the same information if they were interviewed on the same day or if they have the same prior beliefs because of either similar levels of political interest or their partisanship.

One potential cost of correlated votes is that they might decrease the effective sample size. If every pairwise correlation is zero, then every citizen brings novel information to the group forecast; in contrast, if every pairwise correlation is one, then all citizens have the same information, and interviewing one citizen is sufficient to tell us how the whole group would forecast.

[1] The number of correct group members follows a binomial distribution. With plurality voting, the probability of a correct group decision is equal to one minus the probability that fewer than half of the members make a correct decision. In R, the following code replicates the above calculation, assuming odd group sizes: `n<-c(5,25,5,25); p<-c(0.6,0.6,0.7,0.7); 1-pbinom((n+1)/2,n,p)`.