International Journal of Forecasting 30 (2014) 43–54

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Combining forecasts: An application to elections

Andreas Graefe [a,*], J. Scott Armstrong [b,c], Randall J. Jones Jr. [d], Alfred G. Cuzán [e]

[a] *LMU Munich - Department of Communication Science and Media Research, Munich, Germany*

[b] *University of Pennsylvania, Wharton School, Philadelphia, PA, United States*

[c] *Ehrenberg-Bass Institute, University of South Australia, Adelaide, Australia*

[d] *University of Central Oklahoma - Department of Political Science, Edmond, OK, United States*

[e] *University of West Florida - Department of Government, Pensacola, FL, United States*

ARTICLE INFO

Keywords:
Election forecasting
Combining
Prediction markets
Polls
Econometric models
Expert judgment

ABSTRACT

We summarize the literature on the effectiveness of combining forecasts by assessing the conditions under which combining is most valuable. Using data on the six US presidential elections from 1992 to 2012, we report the reductions in error obtained by averaging forecasts within and across four election forecasting methods: poll projections, expert judgment, quantitative models, and the Iowa Electronic Markets. Across the six elections, the resulting combined forecasts were more accurate than any individual component method, on average. The gains in accuracy from combining increased with the numbers of forecasts used, especially when these forecasts were based on different methods and different data, and in situations involving high levels of uncertainty. Such combining yielded error reductions of between 16% and 59%, compared to the average errors of the individual forecasts. This improvement is substantially greater than the 12% reduction in error that had been reported previously for combining forecasts.

1. Introduction

Combining has a rich history, and not only in forecasting. In 1818, Laplace wrote, "in combining the results of these two methods, one can obtain a result whose probability law of error will be more rapidly decreasing" (as cited by Clemen, 1989). In using photographic equipment to combine portraits of people, Galton (1879, p. 135) found that "all composites are better looking than their components, because the averaged portrait of many persons is free from the irregularities that variously blemish the look of each of them". In the field of population biology, Levins (1966) noted that, rather than striving for one master model, it is often better to build several simple models which, among them, use all of the information available, and then average them. Zajonc (1962) summarized the related literature in psychology, which dates from the early 1900s. Note that these early applications of combining all related to estimation problems, rather than forecasting.

In more recent years, researchers have adopted combining as a simple and useful approach to reducing forecast error. Armstrong (2001) reviewed the literature in order to provide an assessment of the gains in accuracy that can be achieved by combining two or more numerical forecasts. Across thirty studies, the average forecast had 12% less error than the typical component forecast. In addition, the combined forecasts were often more accurate than even the most accurate component forecast.

One intuitive explanation as to why combining improves the accuracy is that it enables forecasters to use more information, and to do so in an objective manner. Moreover, bias exists both in the selection of data and in the forecasting methods that are used. Often the bias is unique to the data and the method, so that when various methods using different data are combined to make a forecast, the biases tend to cancel out in the aggregate.

The research interest in combining forecasts has increased since the publication of the frequently-cited

* Corresponding author.
E-mail address: a.graefe@lmu.de (A. Graefe).

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.02.005

paper by Bates and Granger (1969). Numerous studies have demonstrated the value of combining and have tested many alternative proposed methods of weighting the components (for example, based on their historical accuracy), rather than using simple equal-weight averages. However, in an early review of more than two hundred published papers, Clemen (1989) concluded that using equal weights provides a benchmark that is difficult for more sophisticated approaches to beat.

In 2004, we started the www.PollyVote.com project, to test the benefits of combining forecasts of US presidential elections. Forecasts of election outcomes, produced by the following methods, were all collected and processed: polls, prediction markets, experts' judgment, and quantitative models. We expected large gains in forecast accuracy, since the availability of forecasts from such diverse methods and data sets provided ideal conditions for combining (Armstrong, 2001). We had no strong prior evidence as to the relative performances of the various methods, so we decided to combine the forecasts using equal weights. This approach provided additional benefits, including simplicity of calculation and the resulting potential appeal to a broad audience.

In the following sections, we briefly discuss why and how combining works, and outline the conditions under which it is most useful. We then report the results from our combination forecasts for six US presidential elections, three of which were predicted *ex ante*. The results reveal that combining forecasts under ideal conditions yields large gains in accuracy, much larger than those previously estimated by Armstrong (2001).

2. Why combining reduces forecast error

In this section, we explain the terms used to describe the mechanism of combining that is employed in this study, namely to calculate simple averages of forecasts.

2.1. A note on terms: typical error, combined error, bracketing

The error that is derived by averaging the absolute deviation of a set of N numerical forecasts F_i from the actual value A is termed the "typical error":

$$\frac{\sum_{i=1}^{N} |F_i - A|}{N}.$$

Thus, the typical error is the error that one could expect from a random selection of an individual forecast from a given set of forecasts. In mathematical terms, it is similar to the expected value.

By comparison, the "combined error" is the error that is determined by first averaging the N forecasts F_i, and then comparing that average with the outcome A:

$$\left| \frac{\sum_{i=1}^{N} F_i}{N} - A \right|.$$

When one forecast is higher than the actual score that was predicted, and one is lower, "bracketing" occurs (Larrick & Soll, 2006). That is, the value to be predicted lies within the range of a set of forecasts. In this situation, the combined error will invariably be lower than the typical error. When bracketing does not exist, the typical error and the combined error will be of the same magnitude. In that case, combining will not improve the accuracy, but neither will it diminish it.

2.2. An example from the 2012 election

In the 2012 election, President Obama won 52.0% of the two-party popular vote. Several months before the election, Abramowitz's (2012) "time for change" model predicted that Obama would receive 50.6% of the two-party vote for president, which was 1.4 percentage points lower than the actual result. Near the same time, Klarner's (2012) model predicted that Obama would garner 51.3% of the vote, which was 0.7 percentage points too low. Since both models under-predicted the outcome, no bracketing occurred, and hence, the typical error was equal to the combined error: 1.1 percentage points. That is, combining did as well as randomly picking one of the forecasts. In addition, combining did avoid the risk of picking the forecast model that incurred the largest error. However, it also prevented one from picking the most accurate forecast.[1]

Now, consider a situation in which two forecasts lie one on either side of the true value, bracketing it. The 2012 forecast of the Erikson and Wlezien model (2012a) was 52.6%. Thus, the typical error of the two models of Abramowitz and Erikson & Wlezien was 1.0 percentage points. However, the average of the two forecasts (51.6%) missed the true value by only 0.4 percentage points. In this situation, combining the forecasts of the two models reduced the error of the typical individual model by 60%. In addition, the combined forecast was more accurate than either of the individual forecasts.

3. Conditions in which combining is most useful

Combining is applicable to many estimation and forecasting problems. The only exception is when strong prior evidence exists that one method is best *and* the likelihood of bracketing is very low. Armstrong (2001) proposed *ex ante* conditions under which the gains in accuracy that result from combining are expected to be highest: (1) a number of evidence-based forecasts can be obtained; (2) the forecasts draw upon different methods and data; and (3) there is uncertainty about which forecast is most accurate.

3.1. Use of a number of evidence-based forecasts

Accuracy gains that result from combining are most likely to occur when forecasts from many evidence-based

[1] In most real-world forecasting situations, however, it is difficult to identify the most accurate forecast from among a set of forecasts (see Section 3.3).