ARTICLE IN PRESS
JBR-08924; No of Pages 14
Journal of Business Research xxx (2016) xxx–xxx
Contents lists available at ScienceDirect
Journal of Business Research

Testing weighting approaches for forecasting in a Group Wisdom Support System environment☆

Heiko A. von der Gracht [a,1], Ulrich Hommel [b,2], Tobias Prokesch [b,*], Holger Wohlenberg [c,3]

[a] Friedrich-Alexander-University (FAU) Erlangen-Nuremberg, Lange Gasse 20, 90403 Nuremberg, Germany
[b] EBS Business School, Gustav-Stresemann-Ring 3, 65189 Wiesbaden, Germany
[c] Deutsche Börse AG, Mergenthalerallee 61, 65760 Eschborn, Germany

ARTICLE INFO

Article history:
Received 15 October 2014
Received in revised form 21 March 2016
Accepted 22 March 2016
Available online xxxx

Keywords:
Forecasting
Group wisdom
Decision support system
Decision making
Collective intelligence

ABSTRACT

Decision makers usually seek the best possible information to support their decisions. Yet the more experts a decision maker consults, the more divergent opinions he or she might collect. In particular, the approach of attaining an adequate level of information is of crucial importance for many stakeholders such as financial and political institutions as well as sales departments. Inspired by fact that simple heuristics oftentimes outperform complex optimization models, we test and compare several simple forecast-combining methods, including multiple equally weighted approaches, an "imitate-the-successful" heuristic as well as several other weighting approaches (based on self-assessment, knowledge, and hit rate). Forecasts are collected and processed from the crowd using a novel Group Wisdom Support System (GWSS), which provides an entire forecast distribution and information on the consensus evolution over time. We find that the equally weighted triangular forecasts, a simple 1/*N* heuristic, delivers the most accurate results.

1. Introduction and motivation

The combination of forecasts has a long history, which can be traced back to Laplace's idea that one can obtain an overall result whose error is lower by combining the results of two different methods (Clemen, 1989; Cronbach & Meehl, 1955; de Laplace, 1820). In 1966, Levins suggested to build, combine, and average several simple models instead of building just one major and complex model (Levins, 1966). Meanwhile, in machine learning or statistics, bootstrapping, bagging, stacking, and boosting are some frequently cited approaches based on the idea of combining methods (Breiman, 1996; Hibon & Evgeniou, 2005; LeBlanc & Tibshirani, 1996; Schapire, Freund, Bartlett, & Lee, 1998). Subsequent to these seminal works, numerous other studies have analyzed the benefits of combining forecasts by testing different forms of aggregation. Based on an exhaustive literature review covering more than two hundred contributions, Clemen (1989) already concluded in one of his earlier works that the equal-weighting method tends to outperform alternative aggregation schemes. Holden and Peel (1986) even argue that the *Financial Times*' equal-weight forecasts of macroeconomic indicators build the foundation for the formation of expectations with respect to key macroeconomic indicators.

Although this runs counter to the classical ideal of a "true" forecast being made by a single individual, combining individual forecasts from the same event either from experts or from econometric models results in consistently better forecasts than typical single estimates (Armstrong, 2001; Clemen, 1989; Diebold & Pauly, 1990; Timmermann, 2006). Similarly, recent studies have suggested that simple averaging – from portfolio optimization to U.S. presidential elections – yields remarkably good results as, for example, compared to the average errors of the individual forecasts (DeMiguel, Garlappi, & Uppal, 2009; Fildes & Ord, 2002; Graefe, Armstrong, Jones, & Cuzán, 2014).

From a managerial perspective, three factors encourage the combining of forecasts. First, it improves accuracy and decreases the variance of forecasting errors. Second, combining individual forecasts by means of straightforward approaches provides sufficient improvements in accuracy without unduly elevating complexity, thus enabling managers with little experience to apply these methods. Finally, combining individual estimates can be performed at little or no additional cost (Graefe et al., 2014; Mahmoud, 1989). In most cases, combining only a few forecasts suffices to exploit the power of averaging (Clemen & Winkler, 1985; Hogarth, 1978; Johnson, Budescu, & Wallsten, 2001; Sorkin & Dai, 1994). However, Larrick and Soll (2006) demonstrate that under some conditions it is better not to combine forecasts of experts. Similarly, subsequent studies report that the simple average is

☆ The authors thank the valuable comments and feedback on earlier versions of the paper from Andreas Graefe as well as the participants of the 22nd European Conference on Information Systems (ECIS).

* Corresponding author. Tel.: +49 611 7102 1291.
E-mail addresses: heikovdg@gmx.de (H.A. von der Gracht), ulrich.hommel@ebs.edu (U. Hommel), tobias.prokesch@ebs.edu (T. Prokesch), holger.wohlenberg@deutsche-boerse.com (H. Wohlenberg).
[1] Tel.: +49 911 5302 444.
[2] Tel.: +49 611 7102 1241.
[3] Tel.: +49 69 2 11 1 23 00.

http://dx.doi.org/10.1016/j.jbusres.2016.03.043
0148-2963/© 2016 Published by Elsevier Inc.

not always superior (Minson, Liberman, & Ross, 2011; Schultze, Mojzisch, & Schulz-Hardt, 2012).

To contribute to the current academic discussion, our analysis encompasses a variety of combining methods that have been proposed in the literature including trimmed means, performance-based weighting (Armstrong, 2001; Bates & Granger, 1969), and equal weightings (Graefe et al., 2014). Among others, we will provide further evidence that an equally weighted method outperforms a survey-based method, as well as its individual components, the single point forecast and the interval forecast. However, the motivation of our research is not only grounded in these scientific debates but also driven by very concrete demand of financial and political institutions as well as sales departments, which are seeking for new or modified measurement instruments in their daily collection and analysis of market and expert data.

More specifically, to conduct our research, we have created a novel Group Wisdom Support System (GWSS), which has not been previously tested before. The system not only provides a consensus view but also delivers an entire forecast distribution as well as information on the consensus evolution over time. It helps us to test several weighting mechanisms and also allows us to utilize collective intelligence in the sense of "wisdom of crowds" (Surowiecki, 2005). The GWSS collects three point forecasts and combines them using different methods including equality heuristics, an "Imitate-the-successful" heuristic (Boyd & Richerson, 2004) as well as more advanced weighting approaches based on self-assessment, knowledge and hit rate.

2. The value of combining forecasts

One straightforward explanation of why combining forecasts reduces error proneness is that it allows actors to utilize more information. A more analytical explanation of why combining forecasts reduces the forecast error is related to the concept of bracketing (Herzog & Hertwig, 2009). If two forecasts "bracket" the true outcome, then averaging the two estimates performs better than the average of the individual forecasts. By contrast, if two estimates are on the same side of the true outcome, then averaging them will perform equally well as the average forecast (Larrick & Soll, 2006). In situations where estimates made by individuals bracket the true value at a rate of 40% or higher, averaging can outperform even perfect choosing. Fraundorf and Benjamin (2014) reported that averaging outperformed participants responses although the bracketing rate was only 26%. Herzog and Hertwig (2009) revealed that individuals could improve their decision making by thinking in ways that encourage bracketing. In their experimental setup, 50 of 101 participants were asked to revise their initial date estimates of 40 historical events by using a consider-the-opposite strategy. Compared to the participants' first estimates, the average of the initial and the second estimate reduced error by 4.1%. Prokesch, von der Gracht, and Wohlenberg (2015) showed that the bracketing effect is larger than the updating effect and as such of utmost importance for combining forecasts.

The standards and practices described by Armstrong (2001) introduce several principles to improve the accuracy of the combined forecast. Among others, the author suggests to use equal weights unless there is strong evidence to support unequal weights, to use trimmed means, to use the track record to vary the weights, or to use domain knowledge to improve. Since there is no priority among the aforementioned methods, we aim to test whether we can detect an order among the principles. While previous studies have suggested that individual weights deliver superior results (Ashton & Ashton, 1985; Winkler & Makridakis, 1983), Clemen (1989) favors an equally weighted approach since the simple average performs sufficiently well in comparison to more complex methods. The advantageousness of this approach has been confirmed by several later contributions (Armstrong, 2001; Armstrong & Graefe, 2011; Cuzán & Bundrick, 2009; Czerlinski, Gigerenzer, & Goldstein, 1999; Gigerenzer, Hertwig, & Pachur, 2011; Graefe 2015; Yates, McDaniel, & Brown, 1991). Volz and Gigerenzer (2012), for example, state that a simple 1/*N* heuristic tends to outperform mean-variance optimization in situations with high predictive uncertainty. DeMiguel et al. (2009) estimated that in a 50-asset case one would need 500 years of data before a financial optimization model would outperform a simple heuristic such as 1/*N*. In a recent study, Genre, Kenny, Meyler, and Timmermann (2013) analyzed various methods to combine forecasts from the European Central Bank's Survey of Professional Forecasters and concluded that none of the more sophisticated combining methods consistently outperformed the simple average.

Another principle suggests the application of trimmed means. Armstrong recommends the usage of trimmed means if you have at least five forecasts since individual forecasts might have large errors due to miscalculations, errors in data, or misunderstandings. Jose and Winkler (2008) examined whether trimmed means result in more accurate combined forecasts. The researchers analyzed the effects of using different degrees of trimming (e.g., deleting the five highest and five lowest forecasts from the data set before calculating the average, and so on). In addition, they also investigated whether an alternative forms of averaging, the Winsorized mean, was superior to the trimming. Using probability forecast data from the US and European Surveys of Professional Forecasters, Jose, Grushka-Cockayne, and Lichtendahl (2014) investigated that exterior trimming of inflation forecasts from both surveys performs best and better than the linear opinion pool. If the forecasts are overconfident and not very diverse, the researchers propose to use interior trimming. Other recent studies such as Yaniv (1997) and Harries, Yaniv, and Harvey (2004) showed that trimming is generally used as a decision heuristic when individuals aggregate information in case that uncertainty is prevailing.

A further principle suggested by Armstrong (2001) recommends the consideration of domain knowledge, which can be approximated by the forecasters past performance. After analyzing 111 time series from the M-competition, Makridakis (1990) found that methods did better if they were more accurate in previous ex ante forecast tests. Extending the study of Lobo and Nair (1990); Lobo (1991) examined the impact of several weighting methods on the accuracy of analysts' forecasts of company earnings. Among others, the researcher weighted the forecasts by their previous accuracy and discovered that this approach is more accurate than an equal weight approach. Shamseldin, O'Connor, and Liang (1997) detected after analyzing rainfall-runoff predictions in 11 regions that weighting the forecasts on the previous accuracy of the method reduced the mean absolute percentage error (MAPE) by 14.6%. Similarly, Bjørnland, Gerdrup, Jore, Smith, and Anders (2012) found that weighting forecasts on the past performance is superior to the simple average method. However, another possibility to identify domain knowledge is to ask the forecaster for a self-assessment. Usually, the self-assessment is used to determine the overconfidence inherent in a forecaster's forecast. Studies such as Mathews and Diamantopoulous (1990) or Sanders and Ritzman (2001) provide evidence that such judgmental adjustments tend to improve the overall accuracy.

3. Insights on Group Wisdom Support Systems

Most organizations value the ability to improve group decision-making. Ensuing from this strong practical need, researchers studied the scientific relevance of this objective and developed Group Decision Support Systems (GDSS). In general, "a GDSS combines communication, computing power, and decision technologies to support problem formulation and solution in group meetings" (Desanctis and Gallupe 1987, p. 589). The overall aim of a GDSS is to improve the process of group decision making by eliminating communication barriers, offering appropriate tools and systematically steering the timing and content of the discussion (Desanctis & Gallupe, 1987).

Jessup, Connolly, and Galegher (1990) found that anonymity increases participant involvement toward resolving a given problem within a GDSS. By comparing small (three person), medium (nine-person), and large (eighteen-person) groups, Dennis, Valacich, and

Please cite this article as: von der Gracht, H.A., et al., Testing weighting approaches for forecasting in a Group Wisdom Support System environment, *Journal of Business Research* (2016), http://dx.doi.org/10.1016/j.jbusres.2016.03.043