Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 27 (2011) 183–195
international journal of forecasting
www.elsevier.com/locate/ijforecast

Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task

Andreas Graefe[a,*], J. Scott Armstrong[b]

[a] *Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Germany*
[b] *The Wharton School, University of Pennsylvania, Philadelphia, PA, United States*

Abstract

We conducted laboratory experiments for analyzing the accuracy of three structured approaches (nominal groups, Delphi, and prediction markets) relative to traditional face-to-face meetings (FTF). We recruited 227 participants (11 groups per method) who were required to solve a quantitative judgment task that did not involve distributed knowledge. This task consisted of ten factual questions, which required percentage estimates. While we did not find statistically significant differences in accuracy between the four methods overall, the results differed somewhat at the individual question level. Delphi was as accurate as FTF for eight questions and outperformed FTF for two questions. By comparison, prediction markets did not outperform FTF for any of the questions and were inferior for three questions. The relative performances of nominal groups and FTF were mixed and the differences were small. We also compared the results from the three structured approaches to prior individual estimates and staticized groups. The three structured approaches were more accurate than participants' prior individual estimates. Delphi was also more accurate than staticized groups. Nominal groups and prediction markets provided little additional value relative to a simple average of the forecasts. In addition, we examined participants' perceptions of the group and the group process. The participants rated personal communications more favorably than computer-mediated interactions. The group interactions in FTF and nominal groups were perceived as being highly cooperative and effective. Prediction markets were rated least favourably: prediction market participants were least satisfied with the group process and perceived their method as the most difficult.

Keywords: Forecast accuracy; Group decision-making; Method comparison; Satisfaction; Combining

1. Introduction

In situations where a lack of appropriate or available information precludes one from using quantitative methods, it can be helpful to incorporate human judgment in order to improve the forecast. But how does one get the best forecast when aggregating information from a group of people? While organizations most commonly rely on unstructured face-to-face meetings, it is difficult to find evidence to support the use of this strategy (Armstrong, 2006).

* Corresponding author.
E-mail addresses: graefe@kit.edu (A. Graefe), armstrong@wharton.upenn.edu (J.S. Armstrong).

0169-2070/$ - see front matter © 2010 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2010.05.004

The literature suggests that structured approaches such as nominal groups or Delphi provide more accurate forecasts than traditional meetings.

In recent years, there has been increasing interest in prediction markets as an approach for eliciting information from people, and a number of organizations have started to experiment with them. However, to date, we still do not know much about the performance of prediction markets. The available studies are limited and are often of a small scale. In particular, we do not know of any study that has analyzed their performance relative to either meetings or other structured group techniques for aggregating the knowledge in groups. Since the emergence of the field, there has been no meta-analysis that has analyzed the accuracy of prediction markets.

To address this deficiency, we conducted laboratory experiments which compared unstructured meetings, nominal groups, Delphi, and prediction markets. We analyzed the relative accuracies of the four group techniques on a quantitative judgment task and examined the participants' perceptions of their group and their group process.

2. Group judgment tasks

We selected problems that did not involve highly distributed information. This is not intended to suggest that problems involving information which is highly dispersed among the participants are unimportant, but merely that we addressed a limited — but typical — forecasting situation: general problems for which people are unlikely to possess specific knowledge. In our study, the groups had to come up with estimates for a quantitative judgment task. This task consisted of ten factual questions that required percentage estimates. These questions had correct solutions, but all group members could be expected to have some degree of uncertainty about the answers.

The findings from group performance studies are generally task-specific. The critical question arises as to whether or not the tasks used in experimental settings are representative of 'real-world' problems, and thus allow for generalizations to different types of tasks (such as forecasting problems) or participants. For example, Wright and Ayton (1986) provided some evidence that the findings from calibration studies which used general knowledge questions are not applicable to judgmental forecasting problems, since the two task types involve different levels of uncertainty. Furthermore, it is often argued that student participants in laboratory experiments are laymen, whereas the participants in 'real-world' forecasting scenarios are experts who are expected to possess superior knowledge.

In general, for a judgmental forecasting method to perform well, it should be able to aggregate the relevant information from its members efficiently. Thus, a basic precondition for the generalizability of problems used in experimental settings is the involvement of participants in trying to solve them. Factual questions are commonly analyzed by group decision-making, as they have been shown to spark the interest of student participants.

3. Group techniques

We analyzed four group techniques that differ in the amount and structure of the interactions permitted between group members: face-to-face meetings, nominal groups, the Delphi method, and prediction markets.

3.1. Unstructured face-to-face meetings

Unstructured face-to-face meetings (FTF) allow any form of direct interaction between group members. Although meetings are most common for group decision-making in organizations, they have been shown to be subject to many biases and drawbacks. For example, (1) it requires time and effort for a group to maintain itself; (2) groups tend to aim to reach 'speedy decisions' and do not consider all problem dimensions, and thus tend to pursue a limited train of thought, which leads to a 'central tendency effect' or 'groupthink'; (3) less confident group members, or people from lower hierarchy levels, may keep silent about their reservations because of either group pressures for conformity or implied threats of sanctions; (4) dominant personalities tend to exert an excessive amount of influence on the group. For a summary of these issues, see Van de Ven and Delbecq (1971).

In summary, Armstrong (2006) found little evidence to support the use of meetings for forecasting or decision-making. In addition, meetings are expensive both to schedule and to run. In some situations,