Organizational Behavior and Human Decision Processes 120 (2013) 24–36

Contents lists available at SciVerse ScienceDirect

Organizational Behavior and Human Decision Processes

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

On the relative importance of linear model and human judge(s) in combined forecasting

Matthias Seifert [a,*], Allègre L. Hadida [b,*]

[a] Operations & Technology Area, IE Business School, Maria de Molina, 12, 5th Floor, 28006 Madrid, Spain
[b] Judge Business School and Magdalene College, University of Cambridge, Trumpington Street, Cambridge CB2 1AG, UK

ARTICLE INFO

Article history:
Received 22 March 2010
Accepted 23 August 2012
Available online 4 October 2012
Accepted by William Bottom

Keywords:
Judgmental forecasting
Combined forecasts
Bootstrapping models
Music industry

ABSTRACT

When and to what extent should forecasts rely on linear model or human judgment? The judgmental forecasting literature suggests that aggregating model and judge using a simple 50:50 split tends to outperform the two inputs alone. However, current research disregards the important role that the structure of the task, judges' level of expertise, and the number of individuals providing a forecasting judgment may play. Ninety-two music industry professionals and 88 postgraduate students were recruited in a field experiment to predict chart entry positions of pop music singles in the UK and Germany. The results of a lens model analysis show how task structure and domain-specific expertise moderate the relative importance of model and judge. The study also delineates an upper boundary to which aggregating multiple judgments in model-expert combinations adds predictive accuracy. It is suggested that ignoring the characteristics of task and/or judge may lead to suboptimal forecasting performance.

© 2012 Elsevier Inc. All rights reserved.

Introduction

The question of when and to what extent forecasts should rely on analytical models or human judges is fundamental to many organizations and becomes increasingly important as technologies continue to advance. For instance, hedge fund managers have intensified their use of algorithmic trading approaches to determine the timing, price and quantity of trading orders. In some markets, automated trading can lead to faster interpretations of complex business information, enabling analysts to reduce their reaction time to emerging trends. Similarly, IBM announced the first commercial application of their Watson supercomputer to the healthcare industry. Using analytical data mining combined with image and speech recognition software, IBM's technology promises to complement physicians' expertise and help them diagnose and treat cancer patients in a more efficient way. Most recently, 2012 Oscar nominee "Moneyball" opened the "model versus expert" debate to a broader public audience. The movie, which is an adaptation of Michael Lewis's best-selling book, tells the true story of the Oakland Athletics – a baseball team that achieved astounding success on the pitch by relying on a purely mathematical approach to selecting team members (Lewis, 2004).

Yet, failures in the trading algorithms may have disastrous consequences in the financial markets. Cancer patients may be misdiagnosed and may die when relying purely on Watson's analytical output. And although the Oakland As' reached the playoffs in four consecutive years from 2000 to 2003 and won their first playoff in 2006, they have not finished within ten games of the division lead since then – admittedly partly because the team's model-based strategy has now become commonplace in several major league sports, including baseball, where top teams such as the New York Yankees and Boston Red Sox emulated it with much larger budgets. In a newly leveled playing field that is increasingly correcting for the effects of undervalued assets, however, increased stages of sophistication in mathematical modeling may soon yield decreasing returns: "Everyone can look at the same numbers, there are lots of mathematicians for hire, and so secrets are hard to keep" (Cowen & Grier, 2011). These examples show that traders, physicians and sport managers alike need to continuously re-assess when and to what extent they can sensibly base their decisions on model outputs, and when and to what extent they should complement them with their own judgment.

This debate is at the heart of our study. Consistent with prior research in the field of judgmental forecasting, we specifically focus on the value of combining forecasts generated by *linear* models with judgmental predictions. Existing studies have discussed the strengths and weaknesses of using linear models or human judgments on their own (Dawes, Faust, & Meehl, 1989; Lawrence, Goodwin, O'Connor, & Önkal, 2006; Meehl, 1996; Schoemaker, 1993). However, demonstrations of how a combination of both model and judge can lead to higher forecasting accuracy are few and far between (Blattberg & Hoch, 1990; Lawrence, Edmundson,

* Corresponding authors. Fax: +44 (0) 1223 339701 (A.L. Hadida).
E-mail addresses: matthias.seifert@ie.edu (M. Seifert), a.hadida@jbs.cam.ac.uk (A.L. Hadida).

0749-5978/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.obhdp.2012.08.003

& O'Connor, 1986). Among the few existing studies, it has been suggested that a simple 50/50 weighting between model outputs and managerial judgment is likely to outperform either of the two alone (Blattberg & Hoch, 1990). This finding had a profound impact on our understanding of the use of decision support tools and spawned several follow-up studies (for example, Hoch & Schkade, 1996; Sanders & Ritzman, 1995; Stewart, Roebber, & Bosart, 1997). However, none of the subsequent studies to date have attempted to examine the robustness of the proposed 50/50 split in greater depth.

We do so in the present paper. Specifically, we analyze the influence of task structure, domain-specific expertise and aggregated judgments on the effectiveness of combined model-judge(s) forecasts. Our empirical sample draws on real world music industry experts predicting the chart success of upcoming pop music singles. The following section provides an overview of the relevant literature and introduces our main hypotheses. We then present the forecasting task and give details of our empirical setting and data collection methods. The subsequent section provides a comprehensive description of the methods used for conducting the study. The data analysis and results section discuss regression results, with a particular emphasis on their out-of-sample generalizability. In the discussion and conclusions section we summarize key findings, offer alternative explanations of the studied phenomena and outline further research.

Theoretical background and hypotheses development

Model-Individual judge combinations

Mechanical forecasting models, whether they are heuristic or estimated from historical data (Hoch & Schkade, 1996), process information in a consistent, systematic and logical manner (Blattberg & Hoch, 1990). While linear models tend to generate fewer errors than human judges, those that occur are also more likely to be large and to lead to leptokurtic distributions[1] caused, for instance, by the use of inappropriate logical rules (Meehl, 1996; Peters, Hammond, & Summers, 1974). In contrast, judges are proficient in identifying new prediction variables (Blattberg & Hoch, 1990) and in providing subjective assessments of variables that are difficult to measure objectively, such as ethical, moral or aesthetic judgments (Einhorn, 1974). Yet, judges are likely to be biased (Kahneman & Tversky, 1984), influenced by organizational politics (Shapira, 2002), and inconsistent when using information.

Judgmental forecasting research has therefore frequently led to the conclusion that predictions based on model-expert combinations prove to be superior to model outputs and expert judgment alone (e.g., McClish & Powell, 1989; Sanders & Ritzman, 1995; Stewart, 1997; Yaniv & Hogarth, 1993). Individual buyers' predictions of catalog sales and individual brand managers' forecasts of coupon redemption rates provide two specific contexts in which the predictive value of model-judge combinations was investigated (Blattberg & Hoch, 1990). In both settings, the authors not only demonstrate how the aforementioned 50:50 split between model and judge leads to higher forecasting accuracy. They also offer a more refined understanding of the way in which the two forecasting inputs interact. In particular, the complementarity of model and judge seems to be directly linked to judges' ability to process task information in a nonlinear way. Taking this into account, the present study sets out to test a variety of task- and judge-related factors that may influence the robustness of Blattberg and Hoch's (1990) initial findings.

[1] A leptokurtic or "super Gaussian" distribution is defined as a statistical distribution with a positive excess kurtosis.

Task structure

Proponents of Brunswik's (1956) probabilistic functionalism share the view that we live in an objective world in which uncertainty resides only in the minds of decision makers. The degree to which a task is perceived as well- or ill-structured is therefore subjective, and relates to the cognitive abilities of the decision maker.

In this paper, we define "task structure" relative to the level of validity and reliability of the cue contents of a specific task. Ill-structured tasks arise through environmental changes that impact the probabilistic linkages between task input and outcome (Wood, 1986). Our conceptualization shares the commonly adopted view that environmental changes frequently lead to missing information relevant to the task (Fellner, 1961; Frisch & Baron, 1988). In turn, missing information reduces the transparency of means-ends relationships.

This effect is evident in unknown probability distributions linking informational cues and task outcome (Camerer & Weber, 1992), in unknown path sequences, and in the absence of appropriate algorithms for integrating task information (Deng, 1996; Hammond, 1996; Simon, 1977; Steinmann, 1976; Wood, 1986). For instance, when forecasting in disruptive industry environments, a sudden regime change can lead to obsolete historical sales data. Uncertainty may consequently arise regarding the relationship between informational cues and the forecasting event, and, hence, regarding the general type of model that is most appropriate to generate accurate forecasts.

Domain-specific expertise

Expert knowledge is highly organized, domain-specific and not transferable (Feltovich, Prietula, & Ericsson, 2006; Hogarth, 2001). It allows judges to process information more quickly and achieve a higher degree of accuracy than novices (Blattberg & Hoch, 1990; Klein, 2003).

Experts' proficiency in utilizing contextual information is particularly salient when generating forecasts in the context of ill-structured tasks (Armstrong, 1983; Lawrence, Goodwin, O'Connor, & Önkal, 2006; Spence & Brucks, 1997). In fact, research in judgmental forecasting showed that experts were likely to outperform linear models when forecasting ill-structured tasks, thanks to their superior ability to anticipate sudden changes in the data structure (Sanders & Ritzman, 1995). These sudden changes may result from nonlinear relationships among informational cues (Kleinmuntz, 1990; Sanders & Ritzman, 1995; Yaniv & Hogarth, 1993). Hence, experts are likely to select fewer, more diagnostic cues from the contextual knowledge surrounding the task and evaluate these cues more consistently than novices and models (Alexander, 1995; Armstrong, 1983; Spence & Brucks, 1997). Data structures of well-structured tasks are less variable and based on linear relationships between informational cues. The latter enable experts to generate accurate forecasts without having to evaluate contextual information.

We consequently anticipate a positive relationship between the amount of contextual knowledge needed to achieve high forecasting accuracy and the degree to which the forecasting task is ill-structured. Based on the proficiency of experts to utilize contextual knowledge, we also conjecture that the optimal model-judge combination shifts towards a heavier reliance on human judgment when tasks are ill-structured. In contrast, well-structured tasks can be described fairly accurately in terms of linear relationships between informational cues and require proportionally less contextual knowledge to generate forecasts. Optimal model-judge combinations will therefore primarily rely on model components in well-structured task contexts, regardless of domain-specific expertise. In sum, Hypothesis 1 proposes a moderation effect of task structure and domain-specific expertise on the relative importance of linear model outputs and individual human judgments in

Download English Version:

https://daneshyari.com/en/article/888642

Download Persian Version:

https://daneshyari.com/article/888642

Daneshyari.com