European Economic Review 62 (2013) 17–40

Contents lists available at SciVerse ScienceDirect

European Economic Review

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

Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging

Olivier Armantier [a,*], Nicolas Treich [b]

[a] Federal Reserve Bank of New York, United States
[b] Toulouse School of Economics (LERNA, INRA), France

ARTICLE INFO

Article history:
Received 11 November 2011
Accepted 27 March 2013
Available online 26 April 2013

JEL classification:
D81
D84
C91

Keywords:
Belief elicitation
Scoring rules
Experimental economics

ABSTRACT

Proper Scoring Rules (PSRs) are popular incentivized mechanisms to elicit an agent's beliefs. This paper combines theory and experiment to characterize how PSRs bias reported beliefs when (i) the PSR payments are increased, (ii) the agent has a financial stake in the event she is predicting, and (iii) the agent can hedge her prediction by taking an additional action. In contrast with previous literature, the PSR biases are characterized for all PSRs and all risk averse agents. Our results reveal complex distortions of reported beliefs, thereby raising concerns about the ability of PSRs to recover truthful beliefs in general decision-making environments.

1. Introduction

Introduced by statisticians in the 1950s, Proper Scoring Rules (PSRs hereafter) are belief elicitation techniques designed to provide an agent the incentives to report her subjective beliefs in a thoughtful and truthful manner (Savage, 1971). Although it is well known that the most common PSRs are only incentive compatible under expected payoff maximization, the literature does not provide a systematic characterization of the biases – i.e. systematic differences between subjective and reported beliefs – produced by PSRs. The objective of this paper is to better understand the properties of PSRs in general economic environments. More precisely, we characterize the possible PSR biases under three effects: (i) a change in the PSR payments, (ii) the introduction of a financial stake in the event predicted, and (iii) the possibility for the agent to hedge her prediction by taking an additional action. In contrast with the existing literature, our results hold for all PSRs and all risk averse agents. The empirical significance of the biases identified is then tested in a between-subject experiment.

Accurate measurements of probabilistic beliefs have become increasingly important both in practice and in academia. In practice, firms are increasingly turning to their employees to forecast e.g. sales, completion dates, or industry trends.[1] Likewise, numerous websites now collect and report predictions (about e.g. sporting or political events) and opinions (about e.g. consumer products, movies, or restaurants).[2] To be meaningful, the predictions and opinions reported must be informative. Precise belief assessments are also important in academia. In the literature on subjective expectations

* Corresponding author. Tel.: +1 91 76 697350.
E-mail address: olast1@gmail.com (O. Armantier).

[1] Such firms include *Microsoft*, *Google*, *Chevron*, *General Electric*, and *General Motors*.

[2] Such opinion websites include *ePinion*, *Ebay*, *Zagat*, or Amazon. Prediction websites include the *Iowa Electronic Market*, the *Hollywood Stock Exchange*, or *Intrade*.

0014-2921/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.euroecorev.2013.03.008

(Manski, 2004), choice data are complemented with elicited beliefs to explain decisions related to e.g. health, education, labor or retirement. Likewise, experimental economists are increasingly eliciting their subjects' beliefs to understand observed behavior better.

Because they are incentive compatible under expected payoff maximization, PSRs have long been one of the most popular belief elicitation techniques, with applications to accounting, business, education, psychology, finance, and economics.[3] Over the past decade, with the rapid development of prediction markets and opinion websites, there has been renewed interest in PSRs. In particular, several mechanisms based on PSRs have been recently proposed to promote honest feed-back when collecting opinions, reviews or reputation assessments online.[4] In addition, Market Scoring Rules have become a popular mechanism to overcome the liquidity problems that have affected early prediction markets.[5] In short, Market Scoring Rules may be described as follows. A group of agents is sequentially asked to make a prediction about a particular event. Each agent is paid for her prediction according to a PSR, but she also agrees to pay the previous agent for his prediction according to the same PSR. Because of their attractive properties, Market Scoring Rules have been rapidly adopted by several firms operating prediction markets.[6]

It is well known however that PSRs generate biases under risk aversion (Winkler and Murphy, 1970), or under non-expected utility (Offerman et al., 2009). Up to this point, however, the exact nature of the biases induced by PSRs has only been characterized for a few specific PSRs, typically the quadratic scoring rule. Furthermore, PSRs have usually been analyzed in simple decision-making environments in which the agent's wealth varies only with the PSR payments. This paper contributes to the literature by conducting a more systematic characterization of the biases produced by all PSRs in response to three effects relevant to economists.

First, we consider how varying the PSR payment affects reported probabilities. Although experimental economists have long debated how the incentives affect choices (Camerer and Hogarth, 1999), to the best of our knowledge, the problem has not been explicitly addressed for PSRs.[7] We show theoretically that the impact of the PSR payments on reported probabilities is ambiguous. In particular, in contrast with a common belief that PSRs with low payments induce near risk neutral responses, we show that smaller PSR payments can either reduce or reinforce the PSR biases depending on whether the utility function displays increasing or decreasing relative risk aversion.

Second, we consider an environment in which the agent has a financial stake in the event, a situation common in practice. For instance, an agent may be asked to make a prediction about an economic indicator (e.g. the stock market, the inflation rate) or an event (e.g. a flood, his company future sales). Similarly, an agent facing a Market Scoring Rule always has a stake in the event she predicts, as her payment to the previous predictor depends on the outcome of the event. Finally, subjects in public good experiments are often asked to predict the contributions of others.[8] In all those cases, independent of the PSR payments, the agent has a stake, as her income depends on the outcome of the event she is predicting. We generalize Kadane and Winkler (1988) by showing how violations of the "no stake" condition affect the probabilities reported under any PSR. Furthermore, we provide experimental evidence to support this theoretical result.

Third, the agent has the possibility to hedge her prediction by taking an additional action whose payoff also depends on the event. For instance, in the previous examples, the agent predicting an economic indicator may also choose how to diversify her portfolio, while the agent predicting a catastrophic event may also decide on her insurance coverage. Likewise, subjects in public good experiments have to choose their own contributions.[9] As we shall see, because they are not independent, the prediction and the additional action are in general different from what each decision would be if made separately. More specifically, we show theoretically how hedging can create complex distortions in the probabilities reported with a PSR. Most significantly, reported probabilities may not vary when subjective beliefs change. To the best of our knowledge, this is the first characterization of the PSR biases in a model with hedging opportunities. The empirical relevance of this hedging effect is then illustrated in our experiment.

2. Theory

We assume that an agent holds a unique subjective probability $p \in [0, 1]$ that an event occurs. We assume that p cannot be affected by the agent. We consider a standard expected utility framework. We assume that the agent's von Neumann Morgenstern utility function over income, denoted $u(\cdot)$, is thrice differentiable, strictly increasing, and state-independent.

[3] See Camerer (1995), Offerman et al. (2009), or Palfrey and Wang (2009) for references.

[4] E.g. the "Peer Prediction" mechanism of Miller et al. (2005), or the "Collective Revelation" mechanism of Goel et al. (2009). See Chen and Pennock (2010) for a review of similar mechanisms.

[5] See Hanson (2003, 2007), Abramowicz (2007), Ledyard et al. (2009), Healy et al. (2010).

[6] Companies proposing Market Scoring Rules include *Inkling Markets*, *Consensus Point*, *Yoopick*, *Crowdcast*, and *Predictalot*. Companies using Market Scoring Rules internally include *Microsoft*, *General Motors*, *Boeing*, *Motorola*, *and the International Monetary Fund*.

[7] Some have compared financially vs. non-financially incentivized belief elicitation techniques (Rutström and Wilcox, 2009). We extend this analysis by considering variations of strictly positive financial incentives.

[8] See e.g. Croson (2000), Gächter and Renner (2010), or Fischbacher and Gächter (2010).

[9] Although experimental economists have been aware of stakes and hedging opportunities (Palfrey and Wang, 2009; Andersen et al., 2010; Blanco et al., 2010), these issues have often been ignored (Costa-Gomes and Weizsäcker, 2008; Fischbacher and Gächter, 2010). Concerns have also been raised that eliciting beliefs of subjects engaged in a game, even absent any stakes and hedging considerations, may lead to more strategic thinking and therefore affect behavior (Croson, 2000; Rutström and Wilcox, 2009; Gächter and Renner, 2010).