A Market for Trading Forecasts: A Wagering Mechanism

Raja, Aitazaz Ali; Pinson, Pierre; Kazempour, Jalal; Grammatico, Sergio

Published in:
International Journal of Forecasting

Link to article, DOI:
10.1016/j.ijforecast.2023.01.007

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Raja, A. A., Pinson, P., Kazempour, J., & Grammatico, S. (2024). A Market for Trading Forecasts: A Wagering Mechanism. *International Journal of Forecasting*, *40*(1), 142-159. https://doi.org/10.1016/j.ijforecast.2023.01.007

International Journal of Forecasting 40 (2024) 142–159

Contents lists available at ScienceDirect

International Journal of Forecasting

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

A market for trading forecasts: A wagering mechanism[☆]

Aitazaz Ali Raja[a,*], Pierre Pinson[b,c], Jalal Kazempour[d], Sergio Grammatico[a]

[a] Delft Center for Systems and Control, TU Delft, The Netherlands
[b] Imperial College London, Dyson School of Design Engineering, London, United Kingdom
[c] Technical University of Denmark, Department of Technology, Management and Economics, Kongens Lyngby, Denmark
[d] Technical University of Denmark, Department of Wind and Energy Systems, Kongens Lyngby, Denmark

A R T I C L E   I N F O

Keywords:
Mechanism design
Wagering mechanism
Predictive distribution
Elicitation of probabilities
Value of forecast
Scoring rules

A B S T R A C T

In many areas of industry and society, including energy, healthcare, and logistics, agents collect vast amounts of data that are deemed proprietary. These data owners extract predictive information of varying quality and relevance from data depending on quantity, inherent information content, and their own technical expertise. Aggregating these data and heterogeneous predictive skills, which are distributed in terms of ownership, can result in a higher collective value for a prediction task. In this paper, a platform for improving predictions via the implicit pooling of private information in return for possible remuneration is envisioned. Specifically, a wagering-based forecast elicitation market platform has been designed, in which a buyer intending to improve their forecasts posts a prediction task, and sellers respond to it with their forecast reports and wagers. This market delivers an aggregated forecast to the buyer (pre-event) and allocates a payoff to the sellers (post-event) for their contribution. A payoff mechanism is proposed and it is proven that it satisfies several desirable economic properties, including those specific to electronic platforms. Furthermore, the properties of the forecast aggregation operator and scoring rules are discussed in order to emphasize their effect on the sellers' payoff. Finally, numerical examples are provided in order to illustrate the structure and properties of the proposed market platform.

1. Introduction

Forecasting plays a central role in planning and decision-making; as a result, it has always received substantial attention from researchers and practitioners. For a comprehensive review of forecasting and methodological advances, an encyclopedic article by Petropoulos et al. (2022) can be referred to. To produce high-quality predictions, forecasters rely on high-quality data and sophisticated mathematical models. Often, data are collected and held by different owners at different locations; namely, relevant data are distributed both in terms of geography and ownership. The pooling of these distributed data can generate additional value. For example, logistics companies can exchange their data on consumer behavior to improve their forecasting of future inventory demand. Such a forecast improvement by combining or accessing more data from distributed sources is demonstrated in several studies; see Andrade and Bessa (2017) and Messner and Pinson (2019) for an example in energy applications. The general result such that forecasts can

☆ This work was partially supported by the NWO, The Netherlands under research project P2P-TALES (grant n. 647.003.003), the ERC under research project COSMOS (802348), and by COST under the European Network for Game Theory (action CA16228). Pierre Pinson and Jalal Kazempour were additionally supported through the Smart4RES project (European Union's Horizon 2020, No. 864337). The sole responsibility of this publication lies with the authors. The European Union is not responsible for any use that may be made of the information contained therein.

* Corresponding author.
E-mail address: a.a.raja@tudelft.nl (A.A. Raja).

https://doi.org/10.1016/j.ijforecast.2023.01.007

be improved through combination is already well known within the forecasting community. However, in practice, the data owned by firms or individuals are perceived to have a cost when exposed. For businesses, this cost can be incurred in terms of a loss of competitive advantage, for instance, and for individuals it occur through a loss of privacy. Therefore, in order to incentivize generating value from distributed data, the aim of this work is to design platforms for the pooling of predictive information. Such platforms allow for a monetary transfer from a buyer to sellers, who are then compensated for the costs incurred in data collection, processing, modelling, and so forth, without the explicit exposure of their private data. Because of the market context, in this work, the infrastructural costs associated with the data are not considered.

Our work centers around the area of market-based analytics, which can be broadly categorised into *data markets* and *information markets* depending on whether the traded product is raw data or extracted information. Both types of platforms have received increasing attention in the last few decades. In data markets, the key task is the valuation of data based on the contribution of each data seller to a learning task posted by a data buyer (the *client*), typically through a central platform (Agarwal, Dahleh, & Sarkar, 2019; Ghorbani & Zou, 2019). The market platform determines the monetary compensation that corresponds to the data value. Another significant factor in designing data markets is the cost of a seller's privacy loss (Ghosh & Roth, 2011), which plays an important role in determining the value of data; see Spiekermann, Acquisti, Böhme, and Hui (2015) and Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022). For details on data markets, a comprehensive review by Bergemann and Bonatti (2019) can be referred to. Data markets empower data owners (sellers) to have control over the exposure of their private resources and allow buyers to obtain high-quality training data for their learning algorithms and prediction tasks. Despite their huge potential, data markets are not free from limitations and challenges. First, determining the contribution of a particular data set for a buyer is, in principle, a combinatorial problem because of the possible overlap of information among the data sets (Agarwal et al., 2019). Therefore, the computational requirements for data valuation grow exponentially with an increase in the number of sellers and, consequently, the requirements for the evaluation of remuneration. Second, each seller may regard their data privacy with different levels of sensitivity, which makes designing a privacy-preserving mechanism a challenge. Both of these issues can be addressed, to some extent, by so-called information markets.

Information markets (Linde & Stock, 2011) encompass the trade of a much broader category of information items such as news, translations, legal information, and so forth. Taking this into consideration, *prediction markets* gained popularity beyond academic circles (Berg, Nelson, & Rietz, 2008; Wolfers & Zitzewitz, 2006), and, based on dispersed information, generate aggregate forecasts for uncertain future events by utilizing the notion of the "wisdom of crowds". For example, in a prediction market designed to forecast the result of an election, the share price of political candidates indicates the aggregate opinion on the probability of a candidate's win. In contrast to the structure of prediction markets, an information market for the improvement of a buyer's forecast is designed. This improvement offered by the forecasters is remunerated via a mechanism with formal mathematical guarantees concerning *desirable* economic properties such as the balance of a budget, truthfulness, and so forth (Kilgour & Gerchak, 2004). Consequently, in terms of design, this nature of this work is closer to the markets proposed for forecast elicitation with formal guarantees. In these works, typically, the sellers report their beliefs concerning a future event. Then, after the event occurs, the sellers are ranked according to the quality of their forecasts, evaluated by a *scoring rule* (Gneiting & Raftery, 2007; Kilgour & Gerchak, 2004). One approach that differs from contribution-based rewards such as the "winner takes it all" is proposed in Witkowski, Freeman, Vaughan, Pennock, and Krause (2018). It is of interest to note that rewarding the best encompasses many real-world forecasting settings. For example, Netflix offered 1M USD to the team with the best prediction of how users would rate movies (Witkowski et al., 2018). Despite being popular in forecasting competitions, the "winner takes it all" approach ignores the fact that forecasts other than the best one can still provide additional information. Therefore, in line with the idea of pooling the distributed information, mechanisms that aggregate information provided by all sellers are pursued and rewarded according to their quality.

Here, inspiration is taken from the self-financed *wagering* market setup of Lambert et al. (2008), which features a weighted-score mechanism, and it is used as a starting point. In their setup, each player posts a prediction report for an event and wagers a positive amount of money into a common pool. After the occurrence of the event, the wager pool is redistributed among the players according to their relative individual performance. The payoff function is a weighted mixture of strictly proper scoring functions that satisfies several desirable economic properties. Such self-financed mechanisms yield competition in terms of forecast skill. However, since it does not include criteria related to the use of the forecasts, it then ignores their value with regard to a particular application or observer; in other words, there is no external agent that aggregates, utilizes, or is rewarded by the resulting forecast based on the utility it generates. Thus, their setup cannot allow for a utility-driven improvement of the client's forecasts. In this paper, and in contrast to the setup of Lambert et al. (2008), a mechanism is designed and analyzed that considers both the forecast skill of the players and the utility of the forecasts for a decision-maker.

A situation is considered whereby a client (following the terminology of Kilgour & Gerchak, 2004) posts a forecasting task on the market platform, along with the monetary reward they are willing to pay for an improvement in their own belief. In response, the sellers report their forecasts along with their wagers. A central operator then aggregates these forecasts, considering the wagers as corresponding weights, in order to yield the final forecasts that are passed on to the client. It should be noted that, unlike prediction markets, where their mechanism inherently elicits aggregated information in terms of stock prices, the aggregation of forecasts here has to be performed methodically (Winkler, Grushka-Cockayne, Lichtendahl, & Jose, 2019). Therefore, our first goal is to select a suitable aggregation method that reflects players' wagers into the aggregated forecast. Next, a central operator evaluates the quality and contribution of each reported forecast and their corresponding payoffs. Our framework requires a payoff function with a utility component that rewards a contribution to the forecast improvement and a competitive component that evaluates the relative performance of sellers so as to reward or penalize accordingly. As a result, our second goal is to design a collective payoff function, with utility and competitive components that have desirable economic properties.

Our core contribution is to propose a marketplace for aggregate forecast elicitation using a wagering mechanism that is focused on improving the client's utility in terms of an improvement in their forecast. The proposed market model (Section 3.1.3) is general and history-free; it is general in the sense that tasks from any application area can be posted in the form of binary, discrete, or continuous random variables. History-free implies that past data on sellers' performance or market outcome are not utilized; namely, each instance of the market is set up independently. Then, the requirements for the aggregation of forecast reports are provided by utilizing corresponding wagers, and the quantile averaging is compared with the linear pooling method as an example (Section 3.2.1). Finally, a payoff function is designed that rewards the skill of forecasters relative to each other as well as their contribution to the improvement of the utility of the client. It is shown that the proposed payoff function satisfies the desirable economic properties (Section 3.2.3). The remainder of this paper is organized as follows: following the preliminaries covered in Section 2, the proposed market is described in 3. Illustrative examples in Section 4 are used to show and discuss the workings of our approach, while an energy forecasting application and case study are available in Section 5. Conclusions and perspectives for future work are finally gathered in Section 6.

2. Preliminaries

2.1. Forecasting task

Forecasting is a key requisite for decision-making and planning. It is employed in diverse situations such as predicting a candidate's probability of winning an election, projecting the economic condition of a country, businesses forecasting their sales growth for production planning, renewable energy producers making an energy generation forecast for bidding in the market, and so forth. The diverse processes involved in forecasting also translates into the types of forecasting tasks that are encountered by a decision-maker. Broadly speaking, forecasts can be categorized into point forecasts, probabilistic forecasts, and scenarios (Gneiting, 2011; Morales, Conejo, Madsen, Pinson, & Zugno, 2014).

Point forecasts do not communicate the uncertainty associated with the possible outcomes of an event; consequently, an incomplete picture is delivered to a decision-maker. This shortcoming of point forecasts is resolved by probabilistic forecasts, which provide decision-makers with comprehensive information about potential future outcomes; therefore, in this paper, a focus is placed on probabilistic forecasts. A probabilistic forecast consists of a prediction of the probability distribution function (PDF) or of some summary measures of a random variable Y. These summary measures can be quantile forecasts or prediction intervals (Gneiting & Katzfuss, 2014). The market framework proposed in this paper covers all types of probabilistic forecasts, given that the forecast evaluation method satisfies the property of being *strictly proper*. However, in what follows, a focus is placed on forecasting tasks in terms of PDFs for better exposition, and single-category, multi-category, and continuous forecasting tasks are considered. Mathematically, these types of forecasts relate to the forecasting of binary, discrete, and continuous random variables, respectively. Therefore, these cases are sufficient for covering most forecasting tasks that are found in practice. These forecasting tasks will now be described for uncertain events and relevant examples will be provided.

A single-category task covers binary events where the probability of an event happening is forecast. For example, a hedge fund predicting a return from a prospective investment has a single category forecasting task; i.e., whether the quarterly growth of a prospective investment will be greater than $x\%$. In the probabilistic forecasting framework, the task will translate into "the probability of the quarterly growth being greater than $x\%$". A multi-category forecast can be exemplified with a farming company that wants to predict seasonal rainfall in light, moderate, and heavy categories. Here, the forecast takes the form of a discrete probability distribution; for example, the rainfall in the upcoming season being $\{\text{light}, \text{moderate}, \text{heavy}\}$ has the probability distribution $\{0.2, 0.5, 0.3\}$. Even more comprehensive probabilistic information can be obtained by forecasting an event in terms of a continuous probability distribution. For example, a wind energy producer bidding in an electricity market can obtain the whole uncertainty associated with the day-ahead energy generation event by obtaining a forecast in terms of a probability density function (Pinson, 2012; Zhou et al., 2013).

In all of the above three forecasting forms, decision-makers, such as the hedge fund, the farming company, or the energy producer, can also have the in-house capability of forecasting. However, they expect that additional data and expertise can help them improve the quality of their forecasts for better planning and decision-making, which in turn can lead to a higher utility. One way to achieve such a quality improvement is by designing a forecasting market platform where the data and the expertise of expert forecasters can be pooled in return for a competitive reward, depending on the contribution of each expert. When a decision-maker utilizes such a platform for the improvement of a forecast, they expect experts to report their beliefs truthfully instead of gaming the market for higher rewards. Furthermore, the decision-maker requires the improvement offered by the experts to be measurable by formalized criteria. Both the guaranteed truthful reporting and a numeric evaluation of the quality of any probabilistic forecast can be achieved by so-called *scoring rules*.

2.2. Quality, skill, and scoring

At a forecast pooling platform, a scoring rule is required in order to quantify the improvement in the forecast to be used by the decision-maker. Furthermore, it allows us to rank the forecasters and to assign rewards according to their contributions. It should be noted that this assessment is performed in an ex-post sense; namely, after the event has occurred.

Definition 1 (*Scoring Rule*)**.** Let r be a reported probabilistic forecast and ω represent the event observed eventually. Then, a scoring rule $s : (r, \omega) \to \mathbb{R}$ provides a summary measure that assigns a real value for the evaluation of a probabilistic forecast r in view of the realization ω.

In the context of a marketplace in which forecasts are elicited, the role of the scoring rule $s(r, \omega)$ is to encourage players to do their best when it comes to generating valuable predictive information, as well as in incentivizing their honest reporting. These tasks can be achieved by selecting scoring rules that satisfy certain properties. Next, the properties of the scoring rules required in this work are discussed.

2.2.1. Properties of scoring rules

First, we can incentivize the forecasters to report their beliefs truthfully by rewarding them according to a strictly proper scoring rule (Gneiting & Raftery, 2007).

Definition 2 (*Strictly Proper Scoring Rule*)**.** Let a player report a probabilistic forecast r of an uncertain event Y. Let an outcome ω of an event be distributed according to the probability distribution p. Then, a real-valued function $s(\cdot, \omega)$ is called strictly proper when

$$\mathbb{E}_p[s(r, \omega)] < \mathbb{E}_p[s(p, \omega)], \text{ for all } r \neq p.$$

Here, let ϱ be the support of p and f_{PDF} be the probability density function. Then, $\mathbb{E}_p[s(p, \omega)] = \int_{\varrho} s(p, \omega) f_{PDF}(p) dp$.

Later, a strictly proper scoring rule is utilized for our payoff criteria in order to measure the quality of the probabilistic forecasts and reward the players accordingly. There are many such rules reported in the literature; for instance, the Brier score, logarithmic score, quadratic score, and so on (Winkler et al., 1996). In principle, a scoring rule is chosen based on the properties that are suitable for the application. Here, for a strictly proper score rule, two more properties are considered; namely, non-local and sensitivity to distance (Gneiting & Raftery, 2007). These properties consider a complete PDF while ranking, and allocate a higher reward to a forecaster that concentrates the probability more around the realized event. This corresponds to rewarding a higher forecasting skill on a forecaster's behalf. Next, two other properties of scoring rules are described, which are later referred to in order to demonstrate the effect of the choice of scoring rules on the payoff mechanism. This choice is important for implementing our proposed market design in practical scenarios.

Definition 3 (*Non-Local Scoring (Winkler et al., 1996)*)**.** Let the forecasters report a PDF of an event Y and the corresponding outcome ω is observed. Then, a scoring rule is called local if the score depends only on the probability (for a categorical event) or likelihood (for a continuous variable), assigned to ω. Conversely, the rule is not local if it depends on the entire reported PDF.

Definition 4 (*Sensitivity to Distance (Jose, Nau, & Winkler, 2009)*)**.** Let r be a predictive PDF and R the corresponding cumulative distribution function (CDF). Then, a CDF R' is more distant from the value x than R if $R' \neq R, R'(y) \geq R(y)$ for $y \leq x$, and $R'(y) \leq R(y)$ for $y \geq x$. Consequently, a scoring rule s is said to be sensitive to distance if, for a given ω, $s(r, \omega) > s\left(r', \omega\right)$ whenever R' is more distant from R.

In other words, a scoring rule that is sensitive to distance allocates a higher score to the player who assigned a higher probability to the values closer to the observation as compared with the probability assignment to the values farther away (Winkler et al., 1996). Later, in Section 4.4, the properties of locality and sensitivity to distance are numerically illustrated in order to build a better intuition and provide a comparison between scoring rules.

3. Proposed forecast elicitation market design

A setting of a market with a single buyer and multiple sellers is considered for eliciting a probabilistic forecast in the form of a probability distribution of an uncertain future event. In our setting, a buyer is referred to as a client and sellers as players or forecasters. A client posts a forecasting task on the market platform and announces a rate of monetary compensation for any improvement in their own belief. Players with resources and expertise in forecasting the posted task respond by reporting their forecasts along with the wagers. The market then aggregates the received information and delivers it to the client. This aggregated forecast, in turn, is expected to generate a utility for a client in terms of operational improvement. The resulting utility, considering the announced reward rate, is then distributed among the players such that it corresponds to their contribution. It should be noted that the proposed mechanism can generally be used for the elicitation of forecasts of any event that can generate utility, such as the movement of a stock. Next, our market model is formally described, and later the properties of the corresponding payoff distribution function are shown.

3.1. Market model and participants

3.1.1. Client

Let there be a client i_c who is interested in improving their forecast (for example, a generation forecast for their renewable energy asset). A client is parameterized through the following quantities:

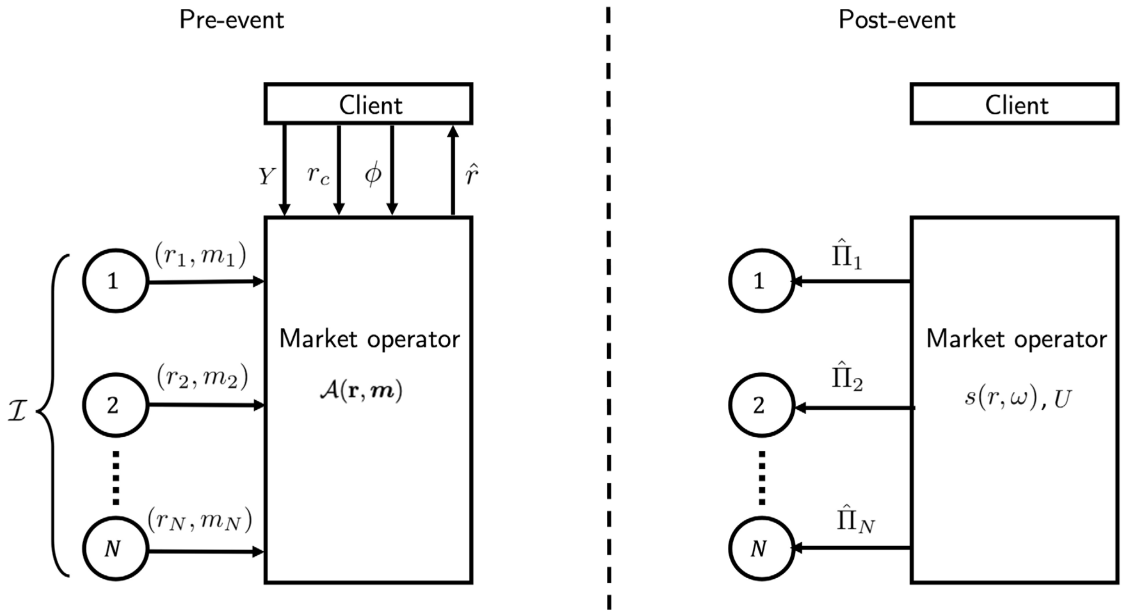


Fig. 1. Market structure showing information flow and pre- and post-event evaluations. The delivery of $\hat{r}$ occurs after all the inputs are received.

- *Forecasting task* Y, an uncertain event that the client wishes to better predict;
- *Forecast report* r_c, a client's own forecast, which is used as a reference for improvement;
- *Reward rate* $\phi > 0$, a monetary value that the client offers for per-unit improvement in the prediction.

A client can post a task Y in the form of a single category forecast (for instance, the probability of energy generation being $[0.4, 0.6]$ per unit), a multi-category forecast (for instance, the discrete probability distribution of energy generation in the intervals $\{[0.4, 0.6], (0.6, 0.8]\}$ per unit), and a continuous forecast (for instance, the probability density function of energy generation). It should be noted that the market design can also accommodate reports in the form of cumulative distribution functions. In what follows, the forecast reports of all three forms are represented by r in order to primarily maintain a focus on the proposed mechanism, which holds for all forms of predictive distributions.

3.1.2. Players

Let $\mathcal{I} = \{1, \ldots, N\}$ be the set of players that are the forecasting experts in the area of a prediction task. A player is parameterized through the following quantities:

- *Forecast report* r_i, a prediction of a forecasting task Y generated using a player i's data resources and expertise; players try to improve r_c in return for a monetary reward;
- *Wager* $m_i > 0$, which accompanies the report r_i and expresses a player i's confidence in their forecast.

A wager is associated with the player's confidence because it decides the level of impact that their prediction has on the resulting forecast. Furthermore, in the proposed payoff function, wagers also influence the reward (penalty) for the players. Here, a penalty implies the partial or complete loss of the amount wagered.

3.1.3. Market operator

A central market operator manages the platform, where a client and the players arrive with respective parameters. This operator is also responsible for maintaining transparency in the market process and is assumed to be honest. The functions of a market operator are:

- evaluation of an aggregated forecast $\hat{r}(\boldsymbol{m}, \boldsymbol{r})$, where $\boldsymbol{r}$ represents a set of predictive distributions $\{r_i\}_{i=1}^N$ posted by the players and $\boldsymbol{m}$ is the vector of corresponding wagers;
- evaluation of the score $s(r_i, \omega)$ of each player $i \in \mathcal{I}$, after observing the outcome ω;
- evaluation of the utility U that corresponds linearly, by assumption, to the improvement in a client's forecast; therefore, in the case of an improvement in the utility $U \propto \phi(s(\hat{r}, \omega) - s(r_c, \omega))$, and is zero otherwise.
- evaluation of the payoff $\hat{\Pi}_i$ of each player $i \in \mathcal{I}$.

Here, after the occurrence of the event, the market operator observes the true outcome ω and evaluates the score $s(r_i, \omega)$ of each player $i \in \mathcal{I}$, which shows how "good" the forecast reported by the player i was. Then, the operator evaluates the utility $U(s(\hat{r}, \omega), s(r_c, \omega), \phi)$ allocated by the client and distributes it among the players that have contributed to the improvement. For transparency, the market operator publicly posts the reward rate, forecast aggregation method, scoring rule, and utility evaluation method, in accordance with the client. The individual predictions posted by the players can be kept private and only an aggregated forecast is delivered to the client. In Fig. 1, the schematic structure of the proposed market with all participants and stages is shown. Note that the allocated utility U depends on the improvement that a client has made and, for the purposes of this work, is treated as an exogenously specified value. Further details on the forecast aggregation methods, the payoff function, and their properties are discussed in what follows.

Remark 1. An important benefit of the proposed market architecture is that the client cannot access the underlying features; instead, they only receive an aggregated forecast. This mitigates a key challenge faced by data markets where sellers are hesitant to release their proprietary data streams since they are freely replicable.

The mechanism design of this market model requires three main components: (*i*) an aggregation operator (to combine forecasts), (*ii*) a scoring rule, and (*iii*) a payoff allocation mechanism. Our goal is to design a history-free mechanism; that is, a mechanism that does not require the past data or reputation of the players in order to compute a solution. This allows our market to be kept general, where clients can post diverse tasks in various forms without an assumption of a repetitive market with a pre-specified task. It should be noted that, in what follows, the arguments from the notations are used and dropped depending on what is necessary. The components of our market mechanism will now be presented and their properties will be discussed.

3.2. Mechanism design

3.2.1. Aggregation operator

After the players have submitted their reports and wagers, in response to the client's forecasting task, the market operator issues a collective forecast $\hat{r}$ using an aggregation operator. Then, the client utilizes the resulting aggregated forecast for the decision-making process, which in turn generates some utility. An improvement in the client's forecast r_c is rewarded at a pre-announced rate ϕ by the client. Therefore, the selection of the forecast aggregation operator constitutes an important part of the mechanism design.

The combining of probabilistic forecasts can be achieved through the weighted averaging of predictive distributions. In this method, a weight assigned to a prediction reflects its relative quality determined by historical data (Knüppel & Krüger, 2022). In other words, the predictions of players are weighted by their historical *performance* and have a corresponding impact on the evaluation of an aggregated forecast. Although logical, such methods are not useful for history-free mechanisms. Therefore, in our proposed mechanism, the performance of a player is associated with their confidence in the reported prediction. Here, the players quantify this confidence through a wagering amount. This enables the assignment of an appropriate weighting to the individual forecasts while combining, which can improve the quality of an aggregated forecast. It also allows our mechanism to penalize (reward) forecasters for low-quality (high-quality) predictions, proportional to their influence on the aggregated forecast via wagers. This penalizing property of the payoff function, referred to as *stimulant*, is referred to in what follows.

Definition 5 (*Aggregation Operator*)**.** An aggregation operator $\mathcal{A} : (\mathbf{r}, \boldsymbol{m}) \rightarrow \hat{r}$ takes a set of predictive reports $\{r_i\}_{i=1}^N$ and a vector of corresponding wagers $\boldsymbol{m} \in \mathbb{R}^N$ as inputs in order to evaluate a combined prediction $\hat{r}$.

Two candidate methods that fulfill the criteria of an aggregation operator are the so-called *linear opinion pool* (LOP) and *quantile averaging* (QA). In terms of distributional forecasts, the linear averaging of probability forecasts can be regarded as vertically combining, and averaging the quantiles can be regarded as horizontally combining (Lichtendahl, Grushka-Cockayne, & Winkler, 2013). Therefore, these two methods can be regarded as two extreme cases in averaging. The first method (LOP) is the most widely used in the literature (Knüppel & Krüger, 2022), as well as in practice. It has several extensions; for instance, the weighted linear opinion pool and the optimally weighted linear opinion pool.

Definition 6 (*Linear Opinion Pool*)**.** Let $\mathcal{I} = \{1, \dots, N\}$ be a set of players. Let r_i be the forecast report of player $i \in \mathcal{I}$ and m_i be the corresponding wager. Then, the LOP is merely an average of all the reports weighted by wagers as $\sum_i \hat{m}_i r_i$ where $\hat{m}_i = \frac{m_i}{\sum_{j \in \mathcal{I}} m_j}$.

For the optimally weighted extension, the weights m_i for all $i \in \mathcal{I}$, are evaluated by setting up an optimization problem that considers past data from the same market. However, even with optimized weights, the LOP suffers from the problem of over-dispersed (under-confident) forecasting, meaning that the aggregate forecast evaluated through the LOP has a higher dispersion than the individual reports (Ranjan & Gneiting, 2010). The authors in Ranjan and Gneiting (2010) proposed a re-calibration method to improve the combined forecast that results from the LOP, where the re-calibration parameters are evaluated by utilizing past data. Thus, this re-calibration method is not suitable for our history-free market mechanism. Next, the quantile averaging is explored, which, interestingly, also corresponds to the Wasserstein barycenter (Agueh & Carlier, 2011) of the reported forecasts.

Definition 7 (*Quantile Averaging*)**.** Let $\mathcal{I} = \{1, \dots, N\}$ be a set of players. For each player $i \in \mathcal{I}$, let r_i be the forecast report in terms of the probability distribution function and R_i be the corresponding cumulative distribution function. Then, the average quantile forecast is given by $\hat{r}_{QA} = \sum_i \hat{m}_i R_i^{-1}$ with $\hat{m}_i$ as in Definition 6.

In Fig. 2, an illustration is presented that shows a comparison of the aggregate forecasts obtained through LOP and QA with equal weights (wagers). Individual forecast reports take the form of Gaussian distributions with different mean and variance values. This illustration provides an idea of how the QA maintains the shape of individual reported forecasts; for instance, here, since individual forecast reports are Gaussian, the aggregated forecast is also Gaussian. This is not the case when using the LOP approach since the resulting aggregate forecast becomes multi-modal. A forecast aggregation based on a QA approach can consequently provide more meaningful aggregated forecasts to decision-makers. Lichtendahl et al.

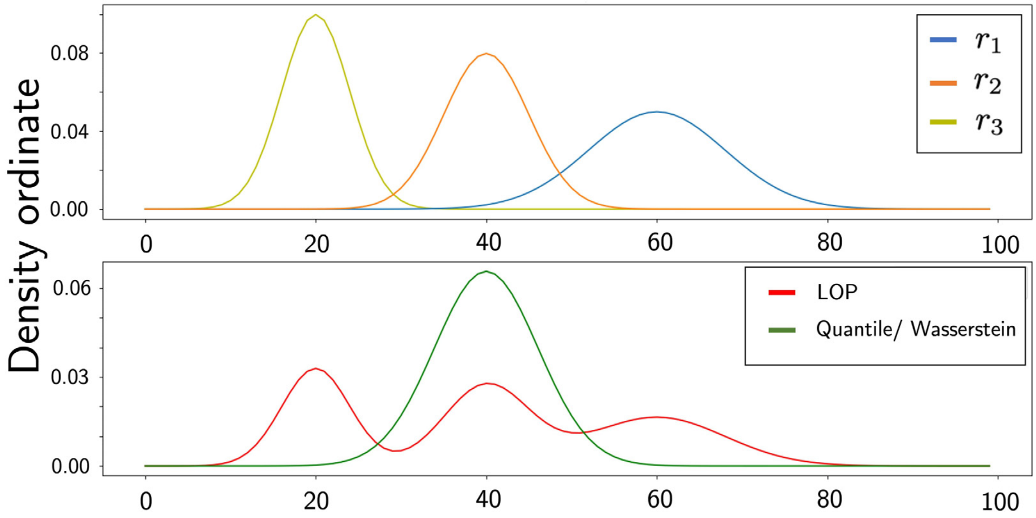


Fig. 2. Comparison of LOP and quantile averaging/Wasserstein barycenter as an aggregation operator.

(2013) extensively covered the useful properties of an aggregated forecast obtained using a QA approach. For instance, an aggregated forecast attained by QA is sharper than that by LOP, and each of its even central moments is less than or equal to those of the LOP (Lichtendahl et al., 2013, Prop. 8). In a memory-free market such as the one being proposed, a prediction that is sharper around the observation can provide more valuable information to the decision-maker and, if probabilistically calibrated, can thus be regarded as being of a higher quality.

It should be noted that the QA can also be interpreted as the report that minimizes the Wasserstein distance $W(\cdot,\cdot)$ from all the forecast reports; that is, $\hat{r} = \min_r \sum_{i=1}^{N} W(r, r_i)$. This then corresponds to the Wasserstein barycenter. Further details on the Wasserstein distance and barycenter can be found in Agueh and Carlier (2011).

Remark 2. The preference for one forecast aggregation method over another is primarily an empirical design choice that is largely application dependent.

3.2.2. Scoring rules

In this subsection, a scoring function $s(r,\omega)$ is specified in order to evaluate the quality of a forecast in an ex-post sense. The continuous ranked probability score (CRPS) is reintroduced, which is a strictly proper score function for the elicitation of a forecast expressed in terms of a probability density function (or, alternatively, a cumulative distribution function). CRPS is non-local and sensitive to distance (see Section 2.2). For single-category and multi-category prediction tasks, scores are presented with similar properties in Appendix A. In order to remain consistent with the relevant scientific literature, scoring rules are defined as negatively-oriented; that is, the lower the better. However, in our payoff function design, which will be presented later, positively oriented scoring will be required. Thus, in what follows, scoring rules may be reoriented for our illustrative examples.

Definition 8 (*Continuous Ranked Probability Score*)**.** For an event of interest x, let r be the forecast report of a given player and let ω be the event that actually occurred; let R denote the cumulative distribution. The continuous ranked probability score (CRPS) is defined as:

$$\text{CRPS}\,(R,\omega) = \int_{-\infty}^{\infty} [R_r(x) - R_\omega(x)]^2\, dx \tag{1}$$

where

$$R_\omega(x) = \begin{cases} 0 & \text{if } x < \omega \\ 1 & \text{if } x \geq \omega \end{cases}$$

In other words, the CRPS provides an assessment of the distance between the forecast report r and the observation ω.

Note that the CRPS can be conveniently reoriented depending on the application. For example, renewable energy production can be normalized to obtain a continuous random variable $P_g \in [0,1]$. Then, the scoring function can be can reoriented by defining $s(r,\omega) = 1 - \text{CRPS}$ and consequently $s(r,\omega) \in [0,1]$. With all the components defined, a wagering-based payoff mechanism and its desired economic properties can now be proposed.

3.2.3. Payoff allocation mechanism

A payoff function is central to the design of a market mechanism as it distributes the pool of wagers $\sum_j m_j$ and the generated utility U among the market players according to their *performance*. Therefore, it is critical for the design of a payoff function that it encourages market participation, on the one hand by clearly reflecting the player's relative contribution and on the other by enabling the delivery of valuable information to the client. The payoff functions are characterized by several desirable properties that can be proven mathematically; for example, budget balance, individual rationality, and so forth.

With regard to the design of a payoff function, inspiration has been taken from Lambert et al. (2008), where the authors present a self-financed wagering mechanism for the elicitation of competitive forecasts. The payoff function in Lambert et al. (2008) rewards the skill of the player relative to the other players by re-distributing the wagers and is shown to satisfy several interesting properties. Such self-financed markets do not involve one particular client with a specific task; as a result, the payoff is only based on the forecast skill of the players and does not involve any utility component. In other words, a player is rewarded for being better than other players regardless of the value or utility of their forecast for a decision-maker. However, our market model in Section 3.1.3 involves a client with a specified task and, therefore, our model involves an external payment associated with the utility of the client. Consequently, a payoff function is required that distributes the utility generated by the forecast; that is, a monetary gain corresponding to an improvement in the client's operational decisions apart from rewarding the forecast skill of the players. In practice, the incentive from the client can implicitly help in improving the forecast quality and in growing the size of the market. For instance, a player who believes their competitors are better informed than them will not enter a market with only a skill payoff, as in Lambert et al. (2008). On the other hand, if the same player believes that their data can provide valuable information and insights to the client in terms of a probabilistic forecast, they will be encouraged to enter our market considering the reward from a utility component. A payoff function shall first be proposed and then its desirable economic properties will be presented.

The payoff function is divided into two parts: one representing the allocation from the wager pool and another from the client's allocated utility. The former evaluates the relative forecasting skill of a player and the latter compensates for their contribution to an improvement in the client's utility U. Let the wager payoff of a player i be

$$\Pi_i(\boldsymbol{r}, \boldsymbol{m}, \omega) := m_i\left(1 + s\left(r_i, \omega\right) - \frac{\sum_j s\left(r_j, \omega\right) m_j}{\sum_j m_j}\right). \quad (2)$$

This term evaluates the relative performance of the players, considering the relative quality of the forecasts and the amounts wagered. It shows that the reward of player i, namely $\Pi_i(\boldsymbol{r}, \boldsymbol{m}, \omega) - m_i$, equals the difference between its performance (confidence and quality) and the average performance of the players. It should be noted that wager payoff can also generate a loss for the players such that they can lose the amount wagered. This is referred to as a penalty to players for posting low-quality forecasts, which plays an important role in showing that our payoff criterion incentivizes truthful reporting by the participants. Now, let us define an indicator $\mathbb{1}_{\{a>b\}}$ that takes a value of 1 if $a > b$ and 0 otherwise. Then, the overall payoff is given by

$$\hat{\Pi}_i = \underbrace{\Pi_i}_{\text{skill component}} + \underbrace{\mathbb{1}_{\{U>0\}}\left(\frac{\tilde{s}\left(r_i, \omega\right) m_i}{\sum_j \tilde{s}\left(r_j, \omega\right) m_j} U\right)}_{\text{utility component}}, \quad (3)$$

where $\tilde{s}(r_i, \omega) = \mathbb{1}_{\{s(r_i,\omega)>\bar{s}\}} s(r_i, \omega)$ and $\bar{s} := s(r_c, \omega)$. Here, the utility component depends on an improvement offered by the player beyond the client's own forecast report r_c. Therefore, in order to be eligible for a share of an allocated utility U, there should first be an improvement in the client's resulting forecast; namely, $U > 0$, and second, the score of player i, $s(r_i, \omega)$ should be greater than the score of the client. Here, the utility payoff of a player is always non-negative, but a skill component can also yield a net loss; that is, $\Pi_i - m_i < 0$ is possible. The possibility of a loss encourages players to compete in improving the forecast by employing better models and acquiring more meaningful data. It should be noted that the client can achieve negative utility as well; that is, the forecast becomes worse than their own prediction. However, again, with a penalty imposed by the wagering part of the payoff function, it is expected that risk-averse players report high-quality forecasts. Next, a brief explanation of some desirable properties of a payoff function is provided.

Desirable properties: The properties are adapted from Lambert et al. (2008) and their explanations are included here in the context of the payoff function in (3).

(i) *Budget-balance*: A mechanism is budget-balanced if the market generates no profit and no loss; that is, $\sum_{i\in\mathcal{I}} \hat{\Pi}_i = \sum_{i\in\mathcal{I}} m_i + U$. In other words, the generated utility and the wager pool must be completely distributed, as a payoff, among the players.

(ii) *Anonymity*: A mechanism satisfies anonymity if the payoff received by a player does not depend on their identity; instead, it depends only on the forecast reports and the realization of an uncertain event. Formally, for any permutation σ of $\mathcal{I}$, the payoff $\hat{\Pi}_i\left((r_i), (m_i), \omega, U\right) = \hat{\Pi}_{\sigma(i)}\left(\left(r_{\sigma^{-1}(i)}\right), \left(m_{\sigma^{-1}(i)}\right), \omega, U\right)$ for all $i \in \mathcal{I}$.

(iii) *Individually rational*: Let the belief of a player $i \in \mathcal{I}$ about an event be p. Then, a mechanism is individually rational if for any wager $m_i > 0$ there exists r_i^* such that an expected profit of a player is non-negative; that is, $\mathbb{E}_p[\hat{\Pi}_i((\boldsymbol{r}_{-i}, r_i^*), \boldsymbol{m}, \omega, U) - m_i] \geq 0$ for any vector of wagers $\boldsymbol{m}_{-i}$ and reports $\boldsymbol{r}_{-i}$.
Individual rationality encourages the participation of players by ensuring a non-negative expected profit according to their beliefs.

(iv) *Sybilproofness*: A truthful mechanism is sybilproof if the players cannot improve their payoff by creating fake identities and copies of their identities. Formally, let the reports $\boldsymbol{r}$ and vectors of wagers $\boldsymbol{m}$ and $\boldsymbol{m}'$ be such that for a subset of players $S \subset \mathcal{I}$ the reports $r_i = r_j$ for $i, j \in S$, the wagers $m_i = m_i'$ for $i \notin S$, and that $\sum_{i\in S} m_i = \sum_{i\in S} m_i'$. Then, the sybilproofness implies that for all $i \notin S$, $\hat{\Pi}_i(\boldsymbol{r}, \boldsymbol{m}, \omega, U) = \hat{\Pi}_i\left(\boldsymbol{r}, \boldsymbol{m}', \omega, U\right)$ and that $\sum_{i\in S} \hat{\Pi}_i(\boldsymbol{r}, \boldsymbol{m}, \omega, U) = \sum_{i\in S} \hat{\Pi}_i\left(\boldsymbol{r}, \boldsymbol{m}', \omega, U\right)$.
It should be noted that the Shapley value, a solution used to evaluate data in a market setting, suffers the drawback of being prone to replication; that is, players can increase their payoff by creating fake copies of themselves (Agarwal et al., 2019). This consideration takes special importance in markets

that deal with forecasts as the data are a freely replicating good.

(v) *Conditionally truthful for players*: A mechanism is conditionally truthful if the player does not have enough information or influence over the payoff function to manipulate it for their own benefit. Therefore, reporting their true belief becomes the best strategy for a risk-averse player.
This definition of conditional truthfulness considers practical situations for the players and the market operation. Truthfulness of a mechanism encourages the players to post their true beliefs on the market platform, thus fulfilling the client's expectation of having access to honest assessments from the experts about an event.

(vi) *Truthful for the client*: A mechanism is truthful for a client, in terms of a reported prediction, if the client's expected payment (allocated utility U) is minimized by reporting their true belief p as their own forecast; that is, $\mathbb{E}_p\left[U(s(\hat{r}, \omega), s(r_c, \omega), \phi)\right] > \mathbb{E}_p\left[U(s(\hat{r}, \omega), s(p, \omega), \phi)\right]$ is satisfied for all $r_c \neq p$. It should be noted that the truthfulness of the client concerns the prediction report r_c and not the reward rate ϕ. With our single-buyer design, it is not possible to elicit their true willingness to pay.

(vii) *Stimulant*: Let a player i's payoff be the sum of skill and utility components; that is, $\pi_i\left(\boldsymbol{r}, (\boldsymbol{m}_{-i}, m_i), \omega, U\right) = \pi_i^s\left(\boldsymbol{r}, (\boldsymbol{m}_{-i}, m_i), \omega\right) + \pi_i^u\left(\boldsymbol{r}, (\boldsymbol{m}_{-i}, m_i), \omega, U\right)$. Let the wager be $m_i' > m_i$; then, this payoff is monotonic if it holds that for the skill component, either

$$0 < \mathbb{E}_p\left[\pi_i^s\left(\boldsymbol{r}, (\boldsymbol{m}_{-i}, m_i), \omega\right) - m_i\right] < \mathbb{E}_p\left[\pi_i^s\left(\boldsymbol{r}, \left(\boldsymbol{m}_{-i}, m_i'\right), \omega\right) - m_i'\right]$$

or

$$0 > \mathbb{E}_p\left[\pi_i^s\left(\boldsymbol{r}, (\boldsymbol{m}_{-i}, m_i), \omega\right) - m_i\right] > \mathbb{E}_p\left[\pi_i^s\left(\boldsymbol{r}, \left(\boldsymbol{m}_{-i}, m_i'\right), \omega\right) - m_i'\right].$$

In other words, a mechanism is monotonic if a player's expected profit, as well as the loss from the skill component, increases by increasing the wager. Now, with regard to the utility factor, let $U > 0$ and $s(r_i, \omega) > \bar{s}$. Then,

$$\pi_i^u\left(\boldsymbol{r}, (\boldsymbol{m}_{-i}, m_i), \omega\right) < \pi_i^u\left(\boldsymbol{r}, \left(\boldsymbol{m}_{-i}, m_i'\right), \omega\right).$$

These properties encourage the players to post higher wagers considering their confidence in their forecasts; as a result, they are referred to as stimulants. Importantly, it also justifies weighting the forecasts by the corresponding wagers while creating an aggregate forecast. It should be noted that, for real-world applications, the market operator can place lower and upper bounds on the amounts of wagers considering the viability of the market.

Now, it is shown that the proposed payoff criterion in (3) satisfies all the desirable properties described above.

Theorem 1 (*Characteristics of Payoff Allocation*)**.** *Let* $s(r, \omega) \in [0, 1]$ *be a strictly proper score function. Then, the payoff function*

$$\hat{\Pi}_i = m_i\left(1 + s\left(r_i, \omega\right) - \frac{\sum_j s\left(r_j, \omega\right) m_j}{\sum_j m_j}\right) + \mathbb{1}_{\{U>0\}}\left(\frac{\tilde{s}\left(r_i, \omega\right) m_i}{\sum_j \tilde{s}\left(r_j, \omega\right) m_j} U\right)$$

is (*i*) *budget-balanced,* (*ii*) *anonymous,* (*iii*) *individually rational,* (*iv*) *sybilproof,* (*v*) *conditionally truthful for players,* (*vi*) *truthful for the client, and* (*vii*) *a stimulant.*

Proof for this theorem has been placed in an Appendix in order to improve flow of this paper.

4. Illustrative examples

In this section, several illustrative examples are outlined in order to provide some information about the proposed market model and to numerically demonstrate the properties of the proposed payoff function in (3). For all the illustrations, a beta distribution is used, with parameters (α, β) as a base predictive density. Then, its parameters are varied in order to simulate potential forecast reports from different players. It has been acknowledged that these reports might not represent a real-world scenario; however, these examples are sufficient for illustrating and discussing the interesting properties of the payoff function.

4.1. Effect of wager amount

Let a client post a prediction task for a random variable $Y \in [0, 1]$ on a market platform along with their own forecast report that has a score of 0.5; that is, $s(r_c, \omega) = 0.5$. In response, let the players $\mathcal{I} = \{1, 2, 3\}$ post their forecast reports as predictive densities that are a random variable Y, as shown in Fig. 3. In real-world cases, it is expected that the reports by expert forecasters are concentrated around nearby values, but here an extreme case is considered in order to emphasize our observations. First, the players' payoff for equal wagers is evaluated, and the wager of player 3 is then increased so as to underline the stimulant property of the payoff function, which is defined in Section 3.2.3. It is supposed that the market operator announces a cap on the wager amount; namely, the maximum value a player can wager, $\bar{m} = 500$. The case of equal wagers in Table 1a shows a loss for player 3, taken from their wager, for posting a sharp predictive density concentrated far from the realized event, $\omega = 0.8$. The corresponding aggregate prediction $\hat{r}_a$, shown in Fig. 3, has a score of 0.867. Here, the score of player 3 is lower than the client's score, and as a result it does not receive any share of the utility payoff. It should be noted that the score from each player is provided by a positively oriented scoring rule (1-CRPS), and the utility of a client is assumed to be specified exogenously. Next, in the case from Table 1b, the wager of player 3 is increased to the maximum acceptable value, which results in an increase in their loss. This implies that showing more confidence by means of a higher wager, albeit in a "bad" forecast, will result in a higher loss. This is an important consequence

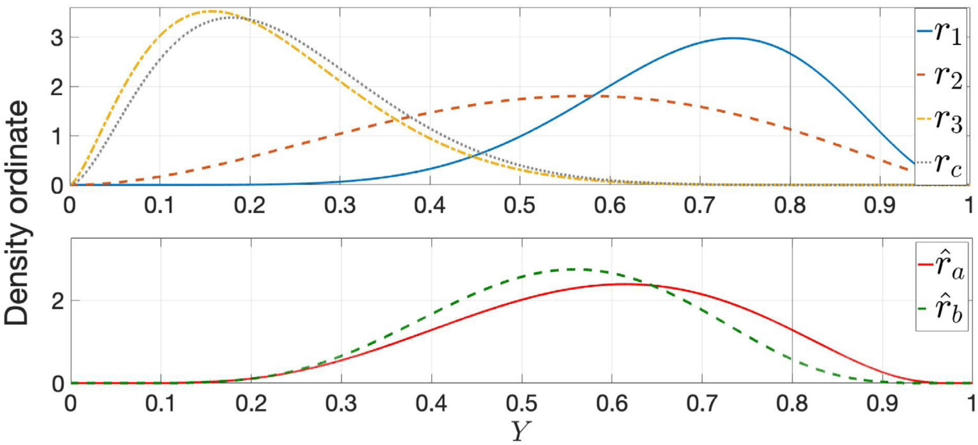


Fig. 3. The top plot shows the reports for density forecasts of a random variable $Y \in [0, 1]$ by market participants, and the bottom plot shows aggregate density forecasts for wagering cases (a) and (b) as in Table 1a and 1b, respectively. The vertical line is at the realization, $\omega = 0.8$.

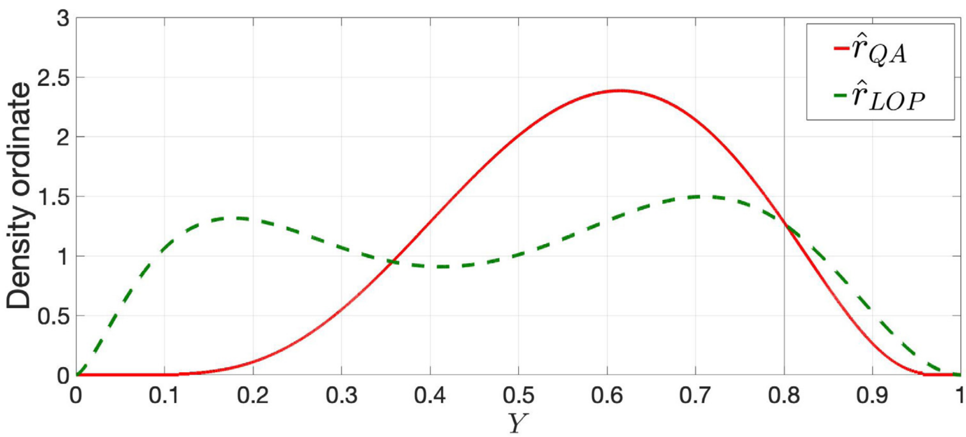


Fig. 4. Plots of aggregate predictive densities obtained through quantile averaging $\hat{r}_{QA}$ and linear opinion pooling $\hat{r}_{LOP}$ in an equal wagering case.

Table 1
Profit (payoff - wager) evaluation for forecast reports in Fig. 3 and its sensitivity to wagers.

(a) Equal wagers			
Players	1	2	3
Wager	100	100	100
Scores	0.9430	0.8450	0.4830
Profit	546	481.39	−27.40
(b) Different wagers			
Players	1	2	3
Wager	100	100	500
Scores	0.9430	0.8450	0.4830
Profit	552.85	488.24	−41.10

for player 3, though also resulting in a reduced quality in the aggregated prediction $\hat{r}_b$, as shown in Fig. 3, with $s(\hat{r}_b, \omega) = 0.822$. This example illustrates the justification for using wagers as weights in the aggregation method. It also shows how using a wager in line with a player's confidence results in a fair penalty or reward for them.

4.2. Comparison of QA and LOP

In Fig. 4, a comparison of aggregate predictive distributions obtained via quantile averaging $\hat{r}_{QA}$ and linear pooling $\hat{r}_{LOP}$ is provided. It is evident how $\hat{r}_{LOP}$ can be problematic for a decision-maker. The loss of sharpness also translates into lower scores for the linear opinion pool, where $s(\hat{r}_{LOP}, \omega) = 0.817$ compared with $s(\hat{r}_{QA}, \omega) = 0.867$. Furthermore, for commonly used parametric distributions, quantile averaging maintains the shape of the distribution, while linear pooling does not.

4.3. Demonstration of sybilproofness

Now, the property of sybilproofness (see Section 3.2.3) is illustrated, which in truthful mechanisms prevents players from manipulating identities. The sybilproofness of the payoff function is especially important in electronic platforms. Table 2a shows the profit and scores of two players with reported predictive densities r_1 and r_2, as in Fig. 3. Now, let the player 2 create a fake identity

Table 2
Sybilproofness of profit (payoff - wager) in the proposed mechanism.

(a) Real identities		
Players	1	2
Wager	100	100
Scores	0.9430	0.8450
Profit	532.30	467.69

(b) Fake identities			
Players	1	2(a)	2(b)
Wager	100	40	60
Scores	0.9430	0.8450	0.8450
Profit	532.30	187.07	280.61

and appear in the market as 2(a) and 2(b) with different wagers, as reported in Table 2b. It can be seen that, even after identity manipulation, the collective profit for both identities of player 2 remained the same as that of the true identity. Consequently, it does not affect the player 1 as well.

4.4. Sensitivity of scoring rules

Various properties of scoring rules are demonstrated in order to emphasize their effect on the design of a payoff function. Generally, the choice of a scoring rule depends on the application area of the prediction task. Therefore, these illustrations are important for providing useful insights into the practitioners when adopting the proposed mechanism to a particular application. The choice of scoring rules can also affect the willingness of players to participate, and constitute an important part of the design.

4.4.1. Local vs. non-local scoring

Different scoring rules differ in their sensitivity to the variation in prediction quality. For applications where sharp predictions are required because of high stakes, scoring rules with a higher sensitivity may perform better. Let us now compare the sensitivity of CRPS and the log score by varying parameters (α, β) of predictive densities. In order to illustrate these effects across the variation in a single parameter α, the mean of the densities is fixed and β is then evaluated as $\beta = \frac{\alpha(1-\text{mean})}{\text{mean}}$. It should be noted that in the parametric case, the variation in parameters simulates the varying quality or features utilized to construct the predictive densities. In Fig. 5(b), the predictive beta distributions for different values of α and the corresponding CRPS and log scores are shown. As the log score depends only on the realization ω, it has a considerable variation for given predictive densities. In contrast, CRPS takes complete information into account and thus varies slightly with the slight change in densities. The scoring rules are selected essentially by considering the nature of the prediction task at hand. It should be noted that our results hold for all strictly proper scoring rules, including the normalized log score.

4.4.2. Sensitivity to distance

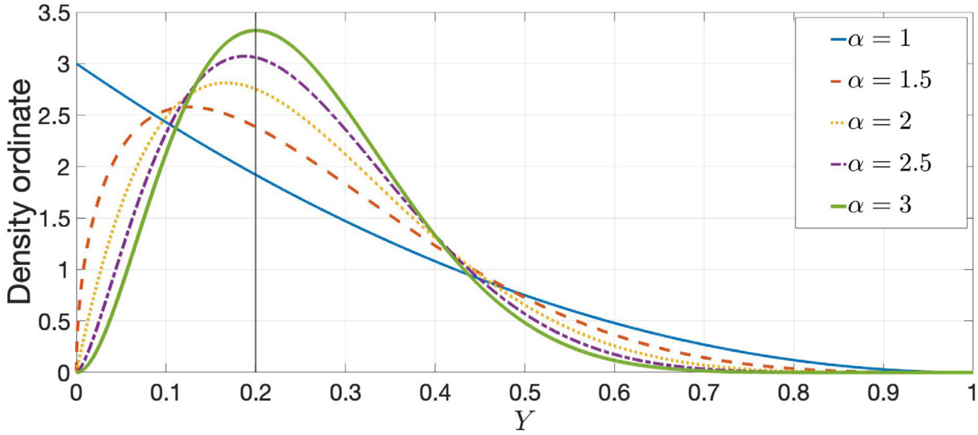
In this example, the impact of the scoring rules' sensitivity to distance is illustrated (see Definition 4). Let the three forecasters E_1, E_2 and E_3 provide normalized multi-category probabilistic forecasts of the energy generation y of a wind producer for intervals $\{[0-0.2], (0.2-0.4], (0.4-0.6], (0.6-0.8], (0.8-1]\}$ per-unit represented by $\{1, 2, 3, 4, 5\}$. Let the reported probabilistic forecasts of E_1, E_2 and E_3 be $\{0.1, 0.1, 0.6, 0.1, 0.1\}$, $\{0, 0.2, 0.6, 0.2, 0\}$ and $\{0.2, 0, 0.6, 0, 0.2\}$, respectively. It is supposed that the actual wind production is observed in the third interval; namely, $y = 3$. Let us now assess the quality of the forecasts using quadratic and ranked probability scoring (RPS) rules (see Winkler et al., 1996 and Appendix A for mathematical expressions). Here, E_1 receives a quadratic score of 0.8, while E_2 and E_3 receive 0.76. It was first observed that all three forecasters had assigned a probability of 0.6 to the realized value of y. Next, it was noted that E_2 assigned the remaining probability of 0.4 to the intervals 2 and 4, which are adjacent to the realized interval; that is, 3, while E_3 assigned it to the most distant intervals. This probability assignment shows comparatively a better forecasting skill on behalf of E_2. However, their scores are the same, which shows that the quadratic scoring is not sensitive to distance. In comparison, RPS assigns 0.975, 0.98 and 0.96 to the predictions of E_1, E_2 and E_3, respectively. It should be noted that RPS acknowledges the concentration of probability around the observation and assigns the highest score to E_2. Therefore, RPS is sensitive to the distance, which can be important for practitioners while designing a payoff function.

5. Wind energy forecasting: A case study

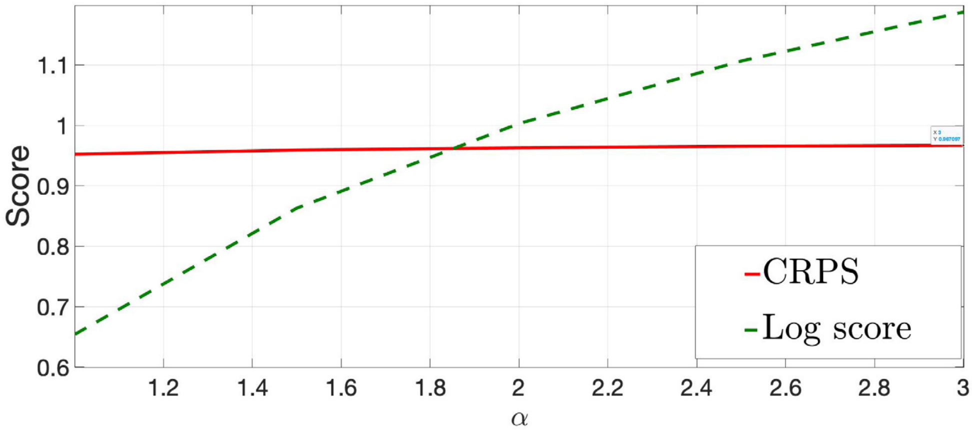
In this section, an energy forecasting application of the proposed market mechanism is presented. Here, forecasters are differentiated based on their forecasting skills and resourcefulness. In the former case (that is, differentiation in terms of forecast skill), the players utilize the same data but different models to construct predictive densities, and vice versa in the latter case (that is, differentiation in terms of access to data). This differentiation allows an important feature of the market to be showcased, which is that it yields a competition for both resourcefulness (data) and forecasting skill among the players. The aim of this case study is then to illustrate how the compensation is allocated by our market mechanism to the various players based on their private information and skills. Elicited forecasts are eventually aggregated and delivered to the client.

5.1. Simulation setup

Here, a wind energy producer is considered as an example: it wishes to improve its generation forecasting and make more informed bids in an electricity market, thereby avoiding a penalty for causing an imbalance. For this purpose, the energy producer arrives at the wagering-based forecasting market, described in Section 3, as a client. It is assumed that the client submits the task of forecasting the next 24 h of wind energy generation. In response, let the forecasters $\mathcal{I}$ submit the probabilistic forecasts along with their wagers. The market operator evaluates the scores of submitted forecasts on an hourly



(a)



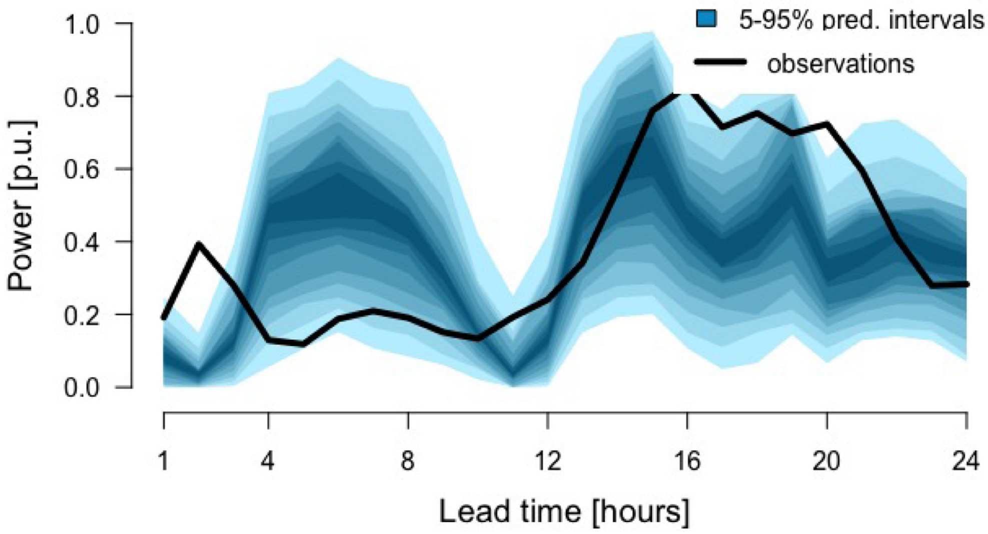
(b)

Fig. 5. (a) Predictive beta distributions with the same mean = 0.25, where for each given α, $\beta = \frac{\alpha(1-\text{mean})}{\text{mean}}$. (b) Comparison of scores assigned to predictive distributions via CRPS and log-score.

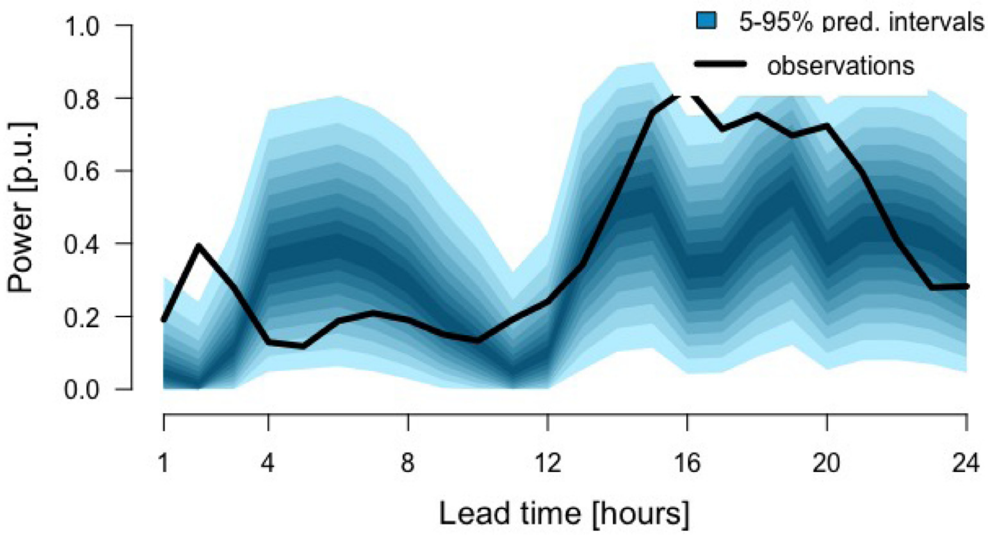
basis and compensates accordingly. For our case study, an open data set from the Global Energy Forecasting Competition 2014, GEFcom2014 (Hong et al., 2016), and an open-source toolkit ProbCast by Browell and Gilbert (2020) are used. The wind power measurements are normalized and thus take values of [0, 1]. For the market setup, a fixed utility U is assumed, which is offered by the client, in order to analyze scores and the share of each player's payoff $\hat{\Pi}_i$ in $\sum_i m_i + U$. It should be noted that, in reality, the compensation provided by the client depends on the operational benefits that they receive through an improvement in their forecast. Next, a simpler case of wind energy forecasting is first presented, with two players evaluating the resulting payoff allocation, as in (3), and later more extensive cases are analyzed.

5.2. Forecasting market with two players

Let the players $\mathcal{I} = \{1, 2\}$ provide a wind energy generation forecast for the next 24 h. Here, it is assumed that both forecasters have the same data but they utilize different models to generate predictive densities for wind energy forecasting. The selection of a particular forecasting model can be seen as a forecasting skill of a player; therefore, the players have different forecasting skills. In this case, player 1 provides their wager m_1 and the forecast report r_1 as a parametric distribution; that is, an inflated beta distribution as proposed by Ospina and Ferrari (2010) generated by using a generalised additive model GAMLSS. In contrast, player 2 utilizes gradient boosted regression trees to generate non-parametric predictive densities and submits the forecast report r_2 along with the wager m_2. Let the market operator announce wager bounds such that $m_1, m_2 \in [10, 100]$. It is assumed that the score of the client's own forecast is constant at 0.5 for all 24 h. Such a low score shows that the client has a low-quality forecast and, consequently, for our data, the players will be eligible for utility payoff at each hour. After receiving the reports, the market operator evaluates an aggregate forecast $\hat{r}$ and delivers it to the wind energy producer (client), who in turn uses it for operational planning. Figs. 6(a) and 6(b) show the reports of player 1 and player 2; that is, r_1 and r_2, respectively. The hourly observations represent the realization ω; that



(a)



(b)

Fig. 6. (a) Wind energy generation forecast reported by player 1 via an inflated beta distribution; that is, r_1. (b) Wind energy generation forecast reported by player 2 via non-parametric predictive density, that is, r_2. Observations represent realization ω.

is, the actual wind energy generation during the corresponding hour. After the forecasting period has passed, the market operator evaluates the score of each player and that of an aggregate forecast. Fig. 7 shows the scores (CRPS) of r_1, r_2 and $\hat{r}$. It has been noted that the aggregate forecast $\hat{r}$ evaluated via quantile averaging, as in Definition 7, depends on the wagers of the players, and Fig. 7 is the case of equal wagers. The difference in the scores of both players is not much since their reported predictive densities follow a similar trend. Though the score rank of players varies at different hours, the parametric forecaster performs slightly better in a cumulative sense for this particular instance of a market. If this variation in score rank is considerable, the aggregate forecast can score better than both players. This fact is illustrated later in our case study. Next, players' total payoffs for 24 h

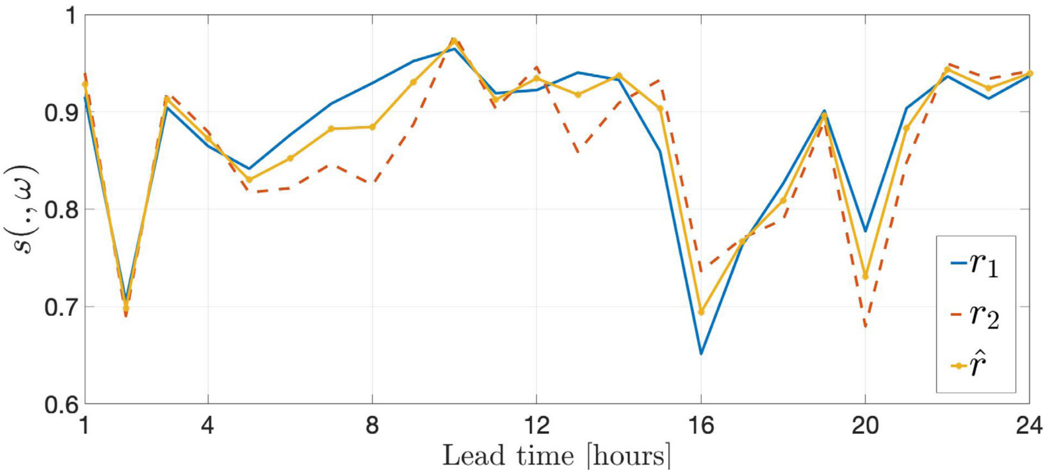


Fig. 7. CRPS of forecasts reported by player 1 (r_1), player 2 (r_2), and an aggregate forecast $\hat{r}$.

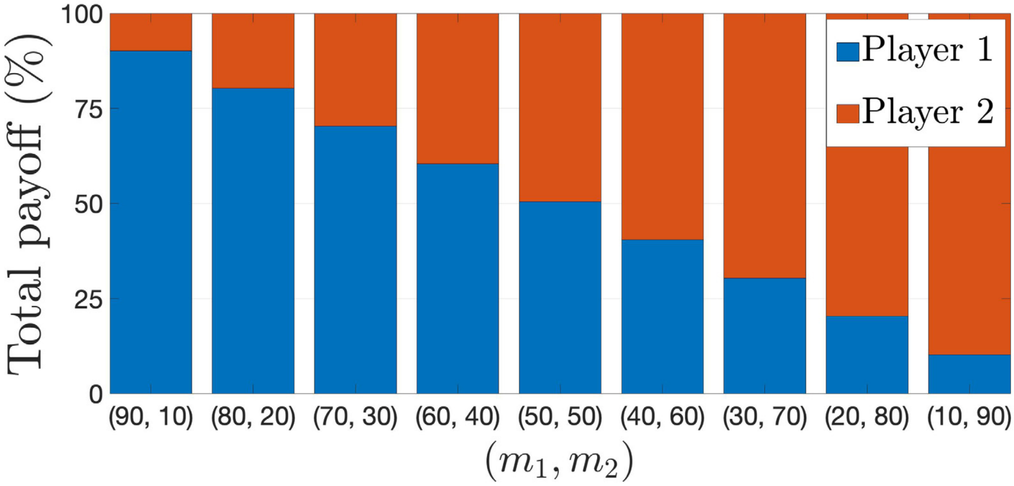


Fig. 8. Players' total payoff of 24 h as a share of money pool $\sum_i m_i + U$ for different wagers.

are shown as a share of the money pool $\sum_i m_i + U$. The payoff, as in (3), also depends on wagers m_i, and in the case of equal wagers, it corresponds directly to the scores. In order to observe the effect of a wager, in Fig. 8, a payoff is plotted across different wager pairs. Since both players offer an improvement and the scores of both players do not differ much, the stimulant property of our payoff function explained in Section 3.2.3 allocates a higher payoff to high-wagering players.

5.3. Forecasting market with four players

Now, it is supposed that two more players join the market and are referred to as player 3 and player 4. It is assumed that these new players have the same forecasting skill; that is, both players utilize the same forecasting method. However, the data held or collected by the players are different. Player 3 holds data on wind forecasts, as a predictor, at a height of 10 m above the ground, whereas player 4 has data on wind forecasts at 100 m above the ground. Wind forecast, being a key predictor, affects the quality of energy generation forecasts. The quality of all four reports is evaluated by CRPS and is presented in Fig. 9 along with the score of an aggregate forecast. In Table 3, the total scores of all forecast reports over a 24-hour period have been reported. Interestingly, in this market instance, the score of the aggregate forecast $s(\hat{r}, \omega)$ is higher than those of the individual forecast reports of all the players.

Table 3
Total score (CRPS) of reported forecasts over a 24-h period.

Report	r_1	r_2	r_3	r_4	$\hat{r}$
Total score	21.0480	20.6978	20.6090	20.2514	21.0074

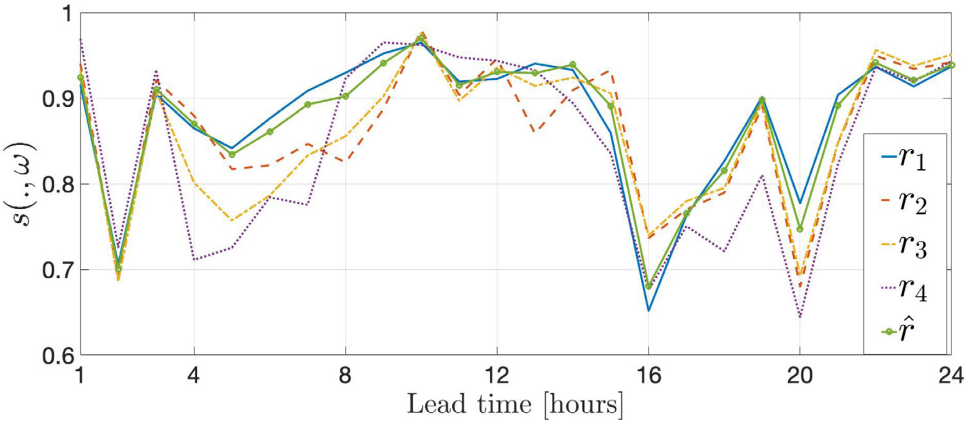


Fig. 9. CRPS of forecasts reported by players (r_1, r_2, r_3 and r_4) and an aggregate forecast $\hat{r}$.

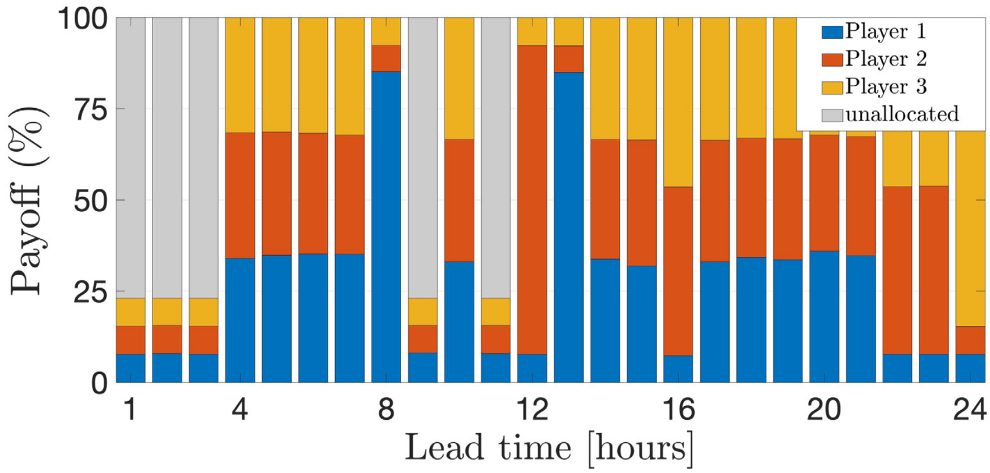


Fig. 10. Players' hourly payoff as a share of the money pool $\sum_i m_i + U$, assuming equal wagers.

In order to analyze the hourly payoff allocation when the client has a forecast report of a reasonable quality, it is assumed that player 4 is the client; that is, $r_c = r_4$, as in (3). Consequently, according to the proposed payoff function in (3), a player becomes eligible for a utility payoff only when it offers an improvement to the client; that is, scores that are higher than the client. Assuming a fixed utility payoff U, players' payoff allocations are presented in Fig. 10. It can be observed that for the first three hours, the score of the client's forecast report (r_4) in Fig. 9 is higher than the players; therefore, the payoff distribution occurs only from the wager pool $\sum_i m_i$. As a fixed utility component U is considered, there remains an unallocated utility payoff component, which is returned back to the client. In contrast, if the utility component depends on the forecast improvement of the client, then $U = 0$ in the case of the first three hours. Next, it was observed that at the 12th hour, only player 2 offered a slight improvement; namely, scores higher than the client (see Fig. 9). As a result, they received the whole offered utility payoff.

6. Conclusions

A marketplace has been designed for the purpose of revealing an aggregate forecast by eliciting truthful individual forecasts from a group of forecasters. In the proposed model, a client with a prediction task calls for forecasts on a market platform and announces a monetary reward for it. The forecasters respond with predictive reports and wagers showing their confidence. The platform aggregates the forecasts and delivers them to the client. Here, the utilized aggregation criteria allow us to make our mechanism a one-shot history-free method that does not account for the forecaster's performance in the past. Next, upon the realization of the event, it allocates payoffs to the forecasters depending on the quality of their forecasts. A payoff function has been proposed with skill and utility components that depend on the relative forecast quality of a forecaster and their contribution to improving the forecast of the client, respectively. It has been shown that the proposed payoff allocation satisfies several desirable economic properties, including budget balance, anonymity, conditional truthfulness, sybilproofness, individual rationality, and stimulant. The simplicity of the scoring-based market design with a wagering mechanism allows it to cater to diverse forecasting tasks with forecasting reports taking forms of discrete to continuous probability distributions.

With the success story of platforms such as NUMERAI (NUMERAI, 2022), there is a real potential for real-world aggregative forecasting marketplaces. Different from current implementations, the mechanism proposed in this paper is designed for the improvement of predictions and provides theoretical guarantees on the monetary compensation, which can encourage and retain the participation of experts. Next, a competition platform has been envisioned in order to test the performance of the proposed market model and the behavior of players in practical scenarios. Such an experimental setup would help us gain further insights for any real-world implementation. Furthermore, our market setup opens several paths for the applied modeling of information eliciting platforms and their analysis. An important step is to design a mechanism for online predictions based on streaming data and, in turn, analyze whether it maintains the economic properties discussed in this paper. Another interesting research avenue is to design models that also value the reputation of forecasters (historical credits).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Scoring rules

Let us present strictly proper scoring rules for single-category and multi-category reporting that are non-local and sensitive to distance (see Section 2.2). A strictly proper scoring rule which is non-local and can be used for eliciting a single-category forecast for binary events is the Brier score.

Definition 9 (*Brier Score*)**.** Let the probability of an occurrence of an event x, reported by a player, be r, and let ω be the actual outcome. Then, the Brier score is given as

$$\mathrm{BS} = (r-\omega)^2 . \tag{A.1}$$

Interestingly, a generalization of the Brier score known as the ranked probability score (RPS), which is also non-local and sensitive to distance, can be used for multi-category forecasting tasks where the reports are in the form of discrete probability distributions.

Definition 10 (*Ranked Probability Score*)**.** Let the multi-category forecasting task have J categories. Let $r(i)$ be the forecast probability of an outcome i, and $\omega(j)$ represents whether the category j has occurred. Then, the ranked probability score is defined as

$$\mathrm{RPS} = \sum_{i=1}^{J} (R(i) - O(i))^2 \tag{A.2}$$

with $R(i) = \sum_{j=1}^{i} r(j)$ and $O(i) = \sum_{j=1}^{i} \omega(j)$.

Appendix B. Proof of Theorem 1

Let us now provide the proof of the properties mentioned in Theorem 1.

1. *Budget balance:* For any vector of reports $\boldsymbol{r}$, wagers $\boldsymbol{m}$, and an outcome ω,

$$\begin{aligned}
\sum_i \hat{\Pi}_i &= \sum_i \Pi_i(\boldsymbol{r}, \boldsymbol{m}, \omega) + \sum_i \frac{\tilde{s}\,(r_i,\omega)\, m_i}{\sum_j \tilde{s}\,(r_j,\omega)\, m_j} U \\
&= \sum_i m_i + \sum_i s\,(r_i,\omega)\, m_i \\
&\quad - \left(\sum_i m_i\right)\left(\frac{\sum_j s\left(r_j,\omega\right) m_j}{\sum_j m_j}\right) \\
&\quad + \sum_i \frac{\tilde{s}\,(r_i,\omega)\, m_i}{\sum_j \tilde{s}\left(r_j,\omega\right) m_j} U \\
&= \sum_i m_i + U.
\end{aligned}$$

2. *Anonymous:* Let σ be any permutation of $\mathcal{I}$. For any $\boldsymbol{r}$, $\boldsymbol{m}$, ω, and i,

$$\begin{aligned}
&\hat{\Pi}_{\sigma(i)} \left(\left(r_{\sigma^{-1}(j)}\right)_{j\in\mathcal{I}}, \left(m_{\sigma^{-1}(j)}\right)_{j\in\mathcal{I}}, \omega, U \right) \\
&= m_{\sigma^{-1}(\sigma(i))} \Bigg(1 + s\left(r_{\sigma^{-1}(\sigma(i))}, \omega\right) \\
&- \frac{\sum_j s\left(r_{\sigma^{-1}(j)},\omega\right) m_{\sigma^{-1}(j)}}{\sum_j m_{\sigma^{-1}(j)}} \\
&+ \frac{s\left(r_{\sigma^{-1}(\sigma(i))},\omega\right)}{\sum_j s\left(r_{\sigma^{-1}(j)},\omega\right) m_{\sigma^{-1}(j)}} U \Bigg) \\
&= m_i \Bigg(1 + s\,(r_i,\omega) - \frac{\sum_j s\left(r_j,\omega\right) m_j}{\sum_j m_j} \\
&+ \frac{s\,(r_i,\omega)}{\sum_j s\left(r_j,\omega\right) m_j} U \Bigg) \\
&= \hat{\Pi}_i \left(\left(r_j\right)_{j\in\mathcal{I}}, \left(m_j\right)_{j\in\mathcal{I}}, \omega, U \right).
\end{aligned}$$

3. *Individually rational:* The skill factor Π_i of the payoff function in (3) is individually rational according to Theorem 1 in Lambert et al. (2008), and the utility factor is always non-negative. As a result, the payoff $\hat{\Pi}_i$ is individually rational; that is, $\mathbb{E}[\hat{\Pi}_i - m_i] \geq 0$.
4. *Sybilproofness:* Let a vector of reports $\boldsymbol{r}$ and vectors of wagers $\boldsymbol{m}$ and $\boldsymbol{m}'$ such that for a subset of players $S \subset \mathcal{I}$ the reports $r_i = r_j$ for $i, j \in S$, the wagers $m_i = m_i'$ for $i \notin S$, and that $\sum_{i\in S} m_i = \sum_{i\in S} m_i'$. Let players $i \in S$ post a common forecast report r;

then, for any $i \notin S$,

$$\hat{\Pi}_i(\boldsymbol{r}, \boldsymbol{m}, \omega, U) = m_i \left(1 + s(r_i, \omega) - \frac{\sum_{j \notin S} s(r_j, \omega) m_j + s(r, \omega) \sum_{j \in S} m_j}{\sum_{j \notin S} m_j + \sum_{j \in S} m_j} + \frac{\tilde{s}(r_i, \omega)}{\sum_{j \notin S} s(r_j, \omega) m_j + s(r, \omega) \sum_{j \in S} m_j} U \right)$$

$$= m_i' \left(1 + s(r_i, \omega) - \frac{\sum_{j \notin S} s(r_j, \omega) m_j' + s(r, \omega) \sum_{j \in S} m_j'}{\sum_{j \notin S} m_j' + \sum_{j \in S} m_j'} + \frac{\tilde{s}(r_i, \omega)}{\sum_{j \notin S} s(r_j, \omega) m_j' + s(r, \omega) \sum_{j \in S} m_j'} U \right)$$

$$= \hat{\Pi}_i(\boldsymbol{r}, \boldsymbol{m}', \omega, U).$$

Additionally, for all $i \in S$

$$\sum_{i \in S} \hat{\Pi}_i(\boldsymbol{r}, \boldsymbol{m}, \omega, U) = \sum_{i \in S} m_i \left(1 + s(r, \omega) - \frac{\sum_{j \notin S} s(r_j, \omega) m_j + s(r, \omega) \sum_{j \in S} m_j}{\sum_{j \notin S} m_j + \sum_{j \in S} m_j} \right) + \sum_{i \in S} \frac{\tilde{s}(r, \omega) m_i}{\sum_{j \notin S} s(r_j, \omega) m_j + s(r, \omega) \sum_{j \in S} m_j} U$$

$$= \left(\sum_{i \in S} m_i \right) \left(1 + s(r, \omega) - \frac{\sum_{j \notin S} s(r_j, \omega) m_j + s(r, \omega) \sum_{j \in S} m_j}{\sum_{j \notin S} m_j + \sum_{j \in S} m_j} \right) + \frac{\tilde{s}(r, \omega) \sum_{i \in S} m_i}{\sum_{j \notin S} s(r_j, \omega) m_j + s(r, \omega) \sum_{j \in S} m_j} U$$

$$= \left(\sum_{i \in S} m_i' \right) \left(1 + s(r, \omega) - \frac{\sum_{j \notin S} s(r_j, \omega) m_j' + s(r, \omega) \sum_{j \in S} m_j'}{\sum_{j \notin S} m_j' + \sum_{j \in S} m_j'} \right) + \frac{\tilde{s}(r, \omega) \sum_{i \in S} m_i'}{\sum_{j \notin S} s(r_j, \omega) m_j' + s(r, \omega) \sum_{j \in S} m_j'} U$$

$$= \sum_{i \in S} \hat{\Pi}_i(\boldsymbol{r}, \boldsymbol{m}', \omega, U).$$

5. *Conditionally truthful for players:* The skill factor Π_i of the payoff function in (3) is truthful according to Theorem 1 in Lambert et al. (2008). Furthermore, for $U > 0$, utility becomes proportional to the strictly proper score function given that $U \propto \phi(s(\hat{r}, \omega) - s(r_c, \omega))$. As a result, players can maximize utility by reporting their true belief $\mathbf{p}$; that is, $\mathbb{E}_p[U(s(\mathcal{A}(\mathbf{p}, \boldsymbol{m}), \omega), s(r_c, \omega), \phi)] > \mathbb{E}_p[U(s(\mathcal{A}(\mathbf{r}, \boldsymbol{m}), \omega), s(p, \omega), \phi)]$ is satisfied for all $\mathbf{r} \neq \mathbf{p}$. Finally, a player does not have enough information and influence on the term $\left(\frac{\tilde{s}(r_i, \omega) m_i}{\sum_j \tilde{s}(r_j, \omega) m_j} \right)$ in (3) to obtain a beneficial arbitrage between skill and utility factors. Therefore, it has been concluded that the payoff $\hat{\Pi}_i$ is conditionally truthful in practical situations.
6. *Truthful for client:* With regard to the design of utility; namely, $U \propto \phi(s(\hat{r}, \omega) - s(r_c, \omega))$, it is proportional to the strictly proper score function. Furthermore, the predictions of forecasters r_i, for all $i \in \mathcal{I}$ are independent of the original forecast report of the client. Writing r_c the client forecast report, it can consequently be seen that the expected payment of the client (allocated utility U) is minimized when the client posts their true belief p; that is,

$$\mathbb{E}_p\left[U(s(\hat{r}, \omega), s(r_c, \omega), \phi)\right] > \mathbb{E}_p\left[U(s(\hat{r}, \omega), s(p, \omega), \phi)\right], \quad \forall r_c \neq p,$$

and

$$p = r_c \implies \mathbb{E}_p\left[U(s(\hat{r}, \omega), s(r_c, \omega), \phi)\right] = \mathbb{E}_p\left[U(s(\hat{r}, \omega), s(p, \omega), \phi)\right].$$

7. *Stimulant:* For a player $i \in \mathcal{I}$, the skill factor Π_i of the payoff function in (3) is monotone according Theorem 1 in Lambert et al. (2008), and the utility factor is proportional to the wager m_i; therefore, the payoff $\hat{\Pi}_i$ is stimulant. ■

References

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, *14*(4), 218–256.

Agarwal, A., Dahleh, M., & Sarkar, T. (2019). A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM conference on economics and computation* (pp. 701–726).

Agueh, M., & Carlier, G. (2011). Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, *43*(2), 904–924.

Andrade, J. R., & Bessa, R. J. (2017). Improving renewable energy forecasting with a grid of numerical weather predictions. *IEEE Transactions on Sustainable Energy*, *8*(4), 1571–1580.

Berg, J. E., Nelson, F. D., & Rietz, T. A. (2008). Prediction market accuracy in the long run. *International Journal of Forecasting*, *24*(2), 285–300.

Bergemann, D., & Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics*, *11*, 85–107.

Browell, J., & Gilbert, C. (2020). ProbCast: Open-source production, evaluation and visualisation of probabilistic forecasts. In *2020 International conference on probabilistic methods applied to power systems* (pp. 1–6). IEEE.

Ghorbani, A., & Zou, J. (2019). Data Shapley: Equitable valuation of data for machine learning. In *International conference on machine learning* (pp. 2242–2251).

Ghosh, A., & Roth, A. (2011). Selling privacy at auction. In *Proceedings of the 12th ACM conference on electronic commerce* (pp. 199–208).

Gneiting, T. (2011). Quantiles as optimal point forecasts. *International Journal of Forecasting*, *27*(2), 197–207.

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*, 125–151.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, *32*(3), 896–913.

Jose, V. R. R., Nau, R. F., & Winkler, R. L. (2009). Sensitivity to distance and baseline distributions in forecast evaluation. *Management Science*, *55*(4), 582–590.

Kilgour, D. M., & Gerchak, Y. (2004). Elicitation of probabilities using competitive scoring rules. *Decision Analysis*, *1*(2), 108–113.

Knüppel, M., & Krüger, F. (2022). Forecast uncertainty, disagreement, and the linear pool. *Journal of Applied Econometrics*, *37*(1), 23–41.

Lambert, N. S., Langford, J., Wortman, J., Chen, Y., Reeves, D., Shoham, Y., et al. (2008). Self-financed wagering mechanisms for forecasting. In *Proceedings of the 9th ACM conference on electronic commerce* (pp. 170–179).

Lichtendahl, K. C., Jr., Grushka-Cockayne, Y., & Winkler, R. L. (2013). Is it better to average probabilities or quantiles? *Management Science*, *59*(7), 1594–1611.

Linde, F., & Stock, W. G. (2011). *Information markets: A strategic guideline for the I-commerce*. Association for Information Science & Technology.

Messner, J. W., & Pinson, P. (2019). Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *International Journal of Forecasting*, *35*(4), 1485–1498.

Morales, J. M., Conejo, A. J., Madsen, H., Pinson, P., & Zugno, M. (2014). Renewable energy sources—Modeling and forecasting. In *Integrating renewables in electricity markets* (pp. 15–56). Springer.

NUMERAI (2022). NUMERAI, the hardest data science tournament on the planet. https://numer.ai/. [Accessed 07 January 2022].

Ospina, R., & Ferrari, S. L. (2010). Inflated beta distributions. *Statistical Papers*, *51*(1), 111–126.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., et al. (2022). Forecasting: Theory and practice. *International Journal of Forecasting*, *38*(3), 705–871.

Pinson, P. (2012). Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *61*(4), 555–576.

Ranjan, R., & Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *72*(1), 71–91.

Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K.-L. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, *25*(2), 161–167.

Winkler, R. L., Grushka-Cockayne, Y., Lichtendahl, K. C., Jr., & Jose, V. R. R. (2019). Probability forecasts and their combination: A research perspective. *Decision Analysis*, *16*(4), 239–260.

Winkler, R. L., Muñoz, J., Cervera, J. L., Bernardo, J. M., Blattenberger, G., Kadane, J. B., et al. (1996). Scoring rules and the evaluation of probabilities. *Test*, *5*(1), 1–60.

Witkowski, J., Freeman, R., Vaughan, J., Pennock, D., & Krause, A. (2018). Incentive-compatible forecasting competitions. In *Proceedings of the AAAI conference on artificial intelligence*: *vol. 32*.

Wolfers, J., & Zitzewitz, E. (2006). Prediction markets in theory and practice. URL https://www.nber.org/papers/w12083.

Zhou, Z., Botterud, A., Wang, J., Bessa, R. J., Keko, H., Sumaili, J., et al. (2013). Application of probabilistic wind power forecasting in electricity markets. *Wind Energy*, *16*(3), 321–338.