ISSN (Print) 0473-453X
Discussion Paper No. 1268
ISSN (Online) 2435-0982

VALUING ALGORITHMS OVER EXPERTS: EVIDENCE FROM A STOCK PRICE FORECASTING EXPERIMENT

Nobuyuki Hanaki
Bolin Mao
Tiffany Tsz Kwan Tse
Wenxin Zhou

December 2024

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Valuing Algorithms Over Experts: Evidence from a Stock Price Forecasting Experiment*

Nobuyuki Hanaki[†] Bolin Mao[‡] Tiffany Tsz Kwan Tse[§] Wenxin Zhou[¶]

December 4, 2024

Abstract

This study examined participants' willingness to pay for stock price forecasts provided by an algorithm, financial experts, and peers. Participants valued algorithmic advice more highly and relied on it as much as expert advice. This preference for algorithms – despite their similar or even lower performance – suggests a shift in perception, particularly among students, toward viewing AI as a reliable and valuable source. However, this "algorithm appreciation" reduced participants' payoffs, as they overpaid for advice that did not sufficiently enhance performance. These findings underscore the need to develop tools and policies that enable individuals to better assess algorithm performance.

Keywords: algorithms, Becker–DeGroot–Marschak mechanism, experts, financial market, forecasting

JEL Classification Number: C90, G1, G4, G17

*We gratefully acknowledge comments and suggestions from John Duffy, Syngjoo Choi, Angela Sutan, Fumio Ohtake, Takeshi Murooka, participants at East Asian Experimental Behavioral Economics Workshop (Seoul), CREST-CEFM Workshop (Paris), and seminars at University of Limassol, Waseda, and Aoyama Gakuin. This project is financially supported by the Joint Usage/Research Center at ISER, Osaka University, and Japan Society for the Promotion of Science KAKENHI Grant Numbers 18K19954, 20H05631, and 23H00055. The experiment reported in this paper was approved by the Research Ethics Committee at the Institute of Social and Economic Research, Osaka University (No. 20231002). The experiment is pre-registered at `aspredicted.org` #150365 (`https://aspredicted.org/smzt-dfqj.pdf`).

†Corresponding author: Institute of Social and Economic Research, Osaka University, Japan and University of Limassol, Cyprus. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

‡Graduate School of Economics, Hitotsubashi University, Japan. E-mail: meta.bolin.mao@gmail.com

§Institute of Social and Economic Research, Osaka University, Japan. E-mail: tiffany.econ@gmail.com

¶A non-academic institution. E-mail: haruka_ivory@outlook.com

1 Introduction

The launch of ChatGPT by OpenAI in November 2022 has revolutionized human-AI interactions. This technology has demonstrated a significant impact on worker productivity, as shown experimentally by Noy and Zhang (2023), and is regarded as having an effect on society comparable to that of the Industrial Revolution. Debates surrounding generative AI technology highlight both its potential benefits and concerns over possible negative impacts across various domains, including work, education, misinformation, and healthcare. See, for example, Capraro et al. (2024) for an interdisciplinary overview as well as Korinek (2024) for an economist's perspective.

The financial advisory industry is no exception. It is undergoing a significant transformation with the rise of artificial intelligence (AI) advisors (OECD, 2019, 2024). Traditional human advisors, while offering personalized services, face challenges such as high fees, behavioral biases (Foerster et al., 2017), conflicts of interest between client profits and employer profits (Hoechle et al., 2018), misconduct (Dimmock et al., 2018; Egan et al., 2019), and reliance on experience-driven decisions (Linnainmaa et al., 2021). In contrast, AI advisors promise lower costs and data-driven recommendations, making them an attractive alternative for investors (D'Acunto et al., 2019). However, despite the growing adoption of AI, relatively little is known about the actual demand for AI advice and optimal pricing strategies for such services.

Recent studies present mixed results regarding preferences for AI versus human advisors in financial decision-making. For instance, Holzmeister et al. (2023) found that clients prefer delegating investment decisions to algorithms, followed by experts with aligned incentives, and finally to experts compensated with fixed fees. Conversely, Germann and Merkle (2023) suggests that participants prioritize returns but exhibit no strong preferences between human fund managers and investment algorithms when it comes to achieving these returns.

When assessing human advisors, investors consider credentials such as certifications and record of their past performance (Sirri and Tufano, 1998; Freer et al., 2023). This raises an important question: how do investors assess AI advisors that lack comparable credentials? To address this question, this paper explores the demand for advice from algorithms versus human experts in the context of stock price forecasting.

The discussions surrounding human-AI relationships have often focused on "algorithm aversion" (Dietvorst et al., 2015), wherein individuals underutilize algorithms or AI after learning that they are imperfect, despite their superior performance compared to human peers. This behavioral trend has been observed in numerous studies, such as by Burton et al. (2020), Chacon et al. (2022), Gill et al. (2024) and Prahl and Van Swol (2017). This has led to inquiries into how to enhance the adoption of AI tools (see, for example, Dietvorst et al., 2018; Filiz et al., 2021; OECD, 2019). Research has examined whether the degree of algorithm aversion is dependent on the task (Castelo et al., 2019), the domain (Himmelstein and Budescu, 2023), or the incentives individuals face (Greiner et al., 2024).

Most existing studies examining algorithm aversion or appreciation (Logg et al., 2019) have relied on what we call the "free advice paradigm."[1] In these

[1] Jussupow et al. (2020) summarizes 29 publications containing 84 distinct experimental studies related to algorithm aversion. These studies employed the free advice paradigm. Mahmud et al. (2022) presents 80 empirical studies related to algorithm aversion. Only two studies, including Gino (2008) and Sutherland et al. (2016), are related to the cost of advice.

studies, participants predict variables such as weight, medical conditions, song rankings, and stock prices. After making an initial prediction, they receive advice (predictions made by algorithms, experts, or peers) at no cost and are then asked to either adopt this advice or adjust their predictions accordingly. Reliance on the advice is measured by the frequency of adopting the provided recommendation or the extent to which the final prediction deviates from the initial one in line with the advice received.

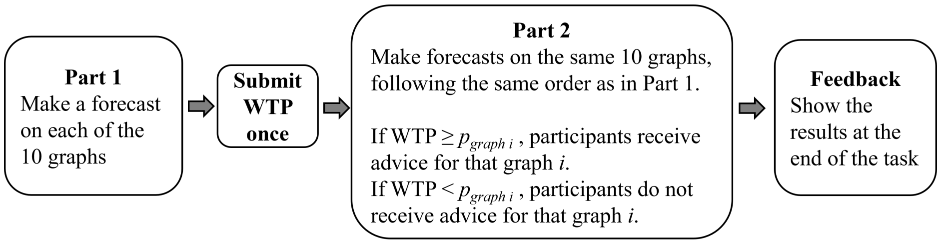
As noted, this advice is offered for free in these studies. However, in many real-life situations – such as investment decisions and medical consultations – advice is not provided without cost. Consequently, before deciding to utilize advice, individuals weigh the cost against the potential benefits. Only after paying for advice can they determine the extent of its use. Thus, rather than solely examining the utilization of free advice, it is essential to understand how much individuals are willing to pay for advice from various information sources and how much they utilize this advice once paid for. To the best of our knowledge, no prior studies have investigated participants' willingness to pay for advice across different sources.

To address this gap in the literature, this paper poses the following research questions: How much are participants in our stock price forecasting experiment willing to pay for forecasts made by an algorithm, experts, and other participants similar to themselves? Furthermore, to what extent do participants utilize or rely on the additional information (forecasts) provided by these sources?

To investigate these questions, we conduct a stock price forecasting experiment akin to those by Bao et al. (2022, 2023) and Tse et al. (2024). In this experiment, participants submit their initial forecasts and have the option to receive additional forecasts (advice) from an algorithm, experts, or similar participants, for which they pay a fee before finalizing their predictions. Employing a between-subject design, each participant receives advice from only one information source, without awareness of the other options available. We utilize the incentive-compatible mechanism proposed by Becker et al. (1964) to elicit participants' willingness to pay for the advice. The tasks are repeated twice to examine how participants' willingness to pay evolves with their experience.

Our findings reveal that participants' willingness to pay is highest for algorithmic advice, followed by expert advice, with the lowest willingness to pay for advice from peers. This pattern persists even after participants complete the task and observe the benefits of receiving advice. Those who gain more from the advice also exhibit a corresponding increase in willingness to pay. However, unbeknownst to participants, the average performance of the algorithm's advice was lower – though not statistically significantly – than advice from other sources in the experiment. As a result, participants improved the accuracy of their forecasts significantly when using advice from experts and peers, but not when using algorithmic advice. Furthermore, while advice from experts and peers generally improved forecasting accuracy, it did not offset the costs incurred: participants' net payoffs decreased significantly on average when they paid for advice versus when they did not, with the highest degree of overpayment associated with the algorithm. In fact, the quality of advice, on average, was no better than the freely available information, i.e., the last observed price. Thus, at least in the context of our experiment, participants would have fared better by forgoing paid advice and instead relying on the last available price for their forecasts.

Figure 1: Flow of Task 1 and Task 2



These results underscore the need for tools and policies that enable individuals to more accurately assess algorithmic performance. From a regulatory perspective, policies promoting transparent disclosures of AI performance metrics, particularly in high-stakes sectors like finance, could help mitigate the risks of overreliance on AI. Educational initiatives in financial and AI literacy could further equip individuals to discern when AI or human expertise is more suitable, potentially reducing financial losses tied to overvaluation of algorithmic advice.

The remainder of the paper is structured as follows: Section 2 presents the experimental design and hypotheses; Section 3 summarizes the results; Section 4 offers a discussion with additional results; and Section 5 concludes the paper.

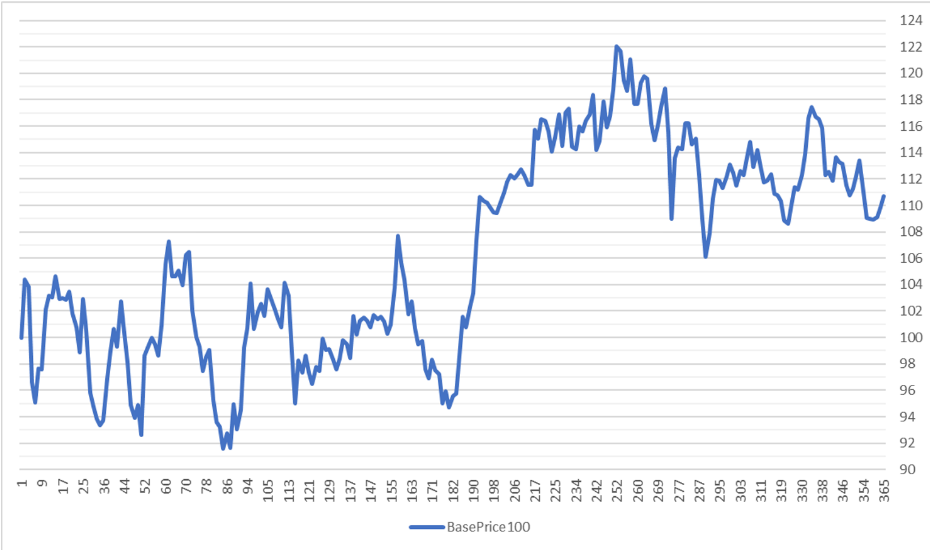
2 Experimental design and hypotheses

2.1 Main tasks

For each treatment, two main tasks were conducted: Task 1 and Task 2. Figure 1 illustrates the flow of these tasks, which are identical in structure. Each task consisted of two parts: Part 1 and Part 2.

In Part 1, participants viewed a series of 10 graphs, each depicting 12

Figure 2: Sample of the graph



months of closing prices for randomly selected stocks from the S&P 500 components. The starting date for each time series was randomly chosen from a range between January 1, 2008, and June 30, 2018, and participants were not informed of either the stock names or the starting dates. To facilitate comparison, each time series was standardized to have a starting price of 100 (see Figure 2 for an example).

Participants were tasked with forecasting the closing price of the stock 30 days after the last price shown on each graph. They entered their forecasts for all 10 graphs, with the display order of the graphs kept consistent across participants. A time limit of 40 seconds was imposed for submitting each forecast. This forecasting task and the graphs utilized in Tasks 1 and 2 adhered to the methodology established in prior forecasting experiments reported by Bao et al. (2022, 2023). During Part 1, participants did not receive feedback on their performance for each graph.

Following Part 1, participants entered the willingness-to-pay (WTP) submission stage, where an incentive-compatible mechanism, as proposed by

Becker et al. (1964, BDM mechanism below), was employed to elicit their WTP for advice on forecasting the stock price graphs presented earlier. Different sources of advice were utilized across treatments, encompassing algorithms, experts, and students, as will be detailed in the subsequent section.

The algorithm-based advice consisted of forecasts generated by algorithms, as outlined in Tse et al. (2024). Expert advice reflected the average forecasts for each graph submitted by Certified Member Analysts (CMAs) of the Securities Analysts Association of Japan, according to Bao et al. (2022). Similarly, student advice comprised the average forecasts for each graph submitted by students from Osaka University, also reported in Bao et al. (2022).

Participants submitted a single willingness-to-pay (WTP) value between 0.0 and 10.0, recorded to one decimal place. They were informed that in Part 2, they would once again make forecasts on the same 10 graphs, presented in the same order as in Part 1. Each participant was notified that the price of advice, denoted as $p_{g,i}$ for each graph i in Part 2, would be randomly determined between 0.1 and 10.0 in increments of 0.1, following a uniform distribution. This randomized advice price varied between participants.

Participants are informed that if their WTP equaled or exceeded $p_{g,i}$ for a given graph i, they would receive the advice for that graph at a cost of $p_{g,i}$. Conversely, if their WTP fell below $p_{g,i}$, they would not receive the advice and would incur no cost. The WTP value established in Task 1 remained constant across all 10 graphs in Part 2. Participants were required to submit this WTP value once in Task 1 and again in Task 2.[2]

[2]It is well known that participants have difficulty understanding the BDM mechanism (Cason and Plott, 2014). To facilitate participants to truthfully report their WTP, in our experiment, we instruct participants to answer the series of questions to determine their WTP: If the price is 0.1, would you like to purchase the advice? If no, your WTP is 0. If yes, ask another question. If the price is 0.2, would you like to purchase the advice? etc.. They are instructed to continue this process starting with $P = 0.1$ and increasing P in 0.1

In Part 2, participants submitted forecasts on the same set of 10 graphs as in Part 1. Prior to each forecast, participants were informed of the price of the advice for that specific graph and whether their WTP allowed them to access the advice. If yes, they were shown both the advice and their previously submitted forecast from Part 1. If not, they were only reminded of their initial forecast from Part 1. Participants then submitted their forecast for each graph in Part 2. No performance feedback was provided during Part 2 for any individual graph.

Upon completing Part 2, participants received comprehensive feedback, including the total points accumulated across both Part 1 and Part 2, as well as the points awarded for their forecasts on each individual graph in both parts. Additionally, participants were reminded of their WTP submitted in Task 1 and the price of advuce associated with each graph. Figure 3 presents a screenshot captured from a participant after completing Task 1.

Participants were explicitly informed that the order of graphs in Part 1 matched that in Part 2; hence, the graph corresponding to question 1 in Part 1 was identical to that in question 11 in Part 2, question 2 in Part 1 corresponded to question 12 in Part 2, and so forth. In this example, the participant was informed that they accrued a total of 232.1 points in Part 1 and 226.9 points in Part 2.

Participants were rewarded based on the accuracy of their forecasts, with the reward for each graph in each part calculated as follows, where $(\cdot)^+$ denotes $\max(\cdot, 0)$:

$$reward = \left(20 - \left|\frac{your\ final\ forecast - realized\ price}{realized\ price}\right| \times 100\right)^+$$

increment until their answer switch from yes to no. See the instruction slide 14 in Online Appendix D.

Figure 3: Screenshot after completion of Task 1 from a participant

Completion of Task 1 and Notification of Results
Thank you for your participation in Task 1.

We will inform you of the results of Task 1.
Total points in Part 1: 132.1 + 100 = 232.1 points
Total points in Part 2: 137.8 + 89.1 = 226.9 points

Please refer to the table below for the details of the gained points.

Click the next button to start Task 2.

Part 1

Questions	Your forecast	Points
1	118.3	10.1
2	140	8.4
3	113	16.8
4	144	15.9
5	100	12
6	143	0
7	133	19.7
8	101	18.6
9	107	17.1
10	123	13.5

Part 2

Questions	WTP	p	Your forecast	Points
11	4.0	1.1	124	4.8
12	4.0	9	140	8.4
13	4.0	5.7	112	17.7
14	4.0	2.3	140	18.8
15	4.0	6.2	107	18.5
16	4.0	9.4	144	0
17	4.0	9.4	133	19.7
18	4.0	5.1	101	18.6
19	4.0	3.8	104	14.4
20	4.0	3.7	119	16.9

Next

If a participant's forecast matched the realized price precisely, they earned 20 points. For every percentage point difference between the forecast and the realized price, 1 point was deducted. Participants received no points if their forecast deviated from the realized price by more than 20%.

To prevent negative payoffs and standardize the initial endowment, each participant was allocated 10 points per graph to cover potential advice costs. The points accrued for each part are represented as:

$$\text{Points gained for each graph in Part 1} = reward_{\text{graph } i} + 10,$$

$$\text{Points gained for each graph in Part 2} = reward_{\text{graph } i} + 10 - p_{\text{graph } i}.$$

The final payoff was determined by summing the points earned across the 10 graphs from a randomly selected part, which could be any of the following:

Table 1: Summary of treatments

Treatments	Algorithm	Expert	Student
No. of participants	106	103	102
No. of sessions	6	5	6
Duration (min)	90	90	90
Avg. payment (JPY)	1884	1907	1921

Part 1 of Task 1, Part 2 of Task 1, Part 1 of Task 2, or Part 2 of Task 2. The exchange rate for the experiment was set at 1 point = 6 JPY.

2.2 Treatments

We employed a between-subjects design to create three distinct treatments, each varying by the source of advice: algorithms, experts, and students. Before participants submitted their WTP for advice from a particular source, they were informed of the method used to generate the advice, though they were not provided with any performance information regarding these sources.[3] In the *algorithm* treatment, participants received advice generated by an algorithm. In the *expert* treatment, advice was provided by experts, while in the *student* treatment, advice was sourced from students.

Table 1 presents a summary of these three treatments. The information given to participants regarding each source of advice was as follows:

Algorithm:

The additional information provided here is a stock price forecast submitted based on the same stock price chart used in this experiment, by an algorithm crafted to predict future stock prices.

[3]Tse et al. (2024) shows that providing such information for the algorithm does not influence participants' reliance on it when they have not experienced the task themselves. Thus, we have decided not provide such information. Thus, WTPs elicited in the Task 1 depend sololy on participants' subjective evaluation.

This algorithm makes the future stock price forecast by learning the historical stock price information, from January 1st, 2000 or from the Initial Public Offering (IPO) day to January 1st, 2020, of 83 target companies rank top in their capital market sectors (i.e. Basic Materials, Consumer Goods, Healthcare, Services, Utilities, Conglomerates, Financial, Industrial Goods, Technology).

Expert:

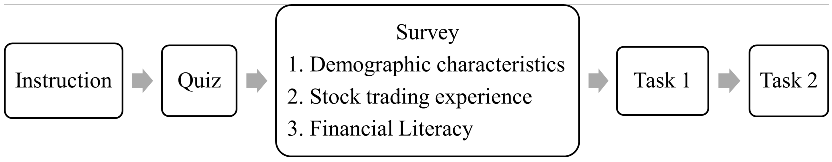
The additional information provided here is the average of stock price predictions submitted by 198 experts (CMAs) who participated in a similar experiment in the past, based on the stock price chart used in this experiment, for the stock prices 30 days ahead.

CMA stands for Certified Member Analyst of the Securities Analysts Association of Japan. The CMA is a qualification granted to those who have taken the prescribed training courses and passed the examinations based on these courses and fulfilled certain requirements, and is a sign of a expert in the fields of finance and investment. The CMA's investment valuation is based on the calculation of the corporate value of an investment and the forecasting of its future value. Therefore, the particularly crucial knowledge can be broadly categorized into three main areas: evaluation of investment data, decision-making on investment policies, and construction and management of portfolios.

Student:

The additional information provided here is the average of stock price predictions submitted by 233 Osaka University students who participated in a similar experiment conducted in the past, based on the stock price chart used

Figure 4: Experiment flow



in this experiment, for the stock prices 30 days ahead.

2.3 Procedure

Figure 4 illustrates the flow of the experiment. Initially, participants read the general instructions individually on a computer screen, which outlined the primary tasks and objectives of the experiment. While participants were informed they would have the opportunity to receive advice, the specific source of this advice was not disclosed at this stage. Detailed instructions are available in the Online Appendix D.

Following the instructions, participants were required to complete a comprehension quiz to ensure they understood the experiment. This quiz consisted of five identical questions across all three treatments. Feedback on each answer was provided immediately after each question. The average quiz scores were 4.491 for the algorithm treatment, 4.379 for the expert treatment, and 4.431 for the student treatment, with no statistically significant differences across treatments ($p = 0.629$).[4] Most participants in all treatments displayed a solid understanding of the experimental rules, with comprehension levels showing no significant variation across treatments.

After the quiz, participants completed a survey consisting of three sec-

[4]Quiz scores across treatments were compared using an Ordinary Least Squares (OLS) regression, with quiz scores regressed on treatment dummies and robust standard errors applied. An F-test compared the estimated dummy coefficients, with p-values used to present results.

tions: demographic characteristics, stock trading experiences, and financial literacy. In the demographic section, participants provided information on their gender, academic grade, and whether they were majoring in a science-related field. For stock trading experience, participants indicated any past involvement in stock investments. Those with prior experience were asked about the duration of their activities, the types of stocks they invested in, their returns, the sources of information they consulted during investment decision-making, and their risk management approaches in stock investments.

In the financial literacy section, participants responded to 12 questions adapted from Fernandes et al. (2014) to assess their level of financial literacy. The mean financial literacy score was 8.019 for both the algorithm and expert treatments, and 8.186 for the student treatment, with no statistically significant differences observed across treatments ($p = 0.816$).[5] (see Online Appendix F for survey details).

Upon completing the survey, participants advanced to the main experimental tasks, Task 1 and Task 2, which followed the same order for all participants. In the algorithm treatment, participants received advice generated by an algorithm; in the expert treatment, advice came from an expert; and in the student treatment, advice was provided by a student.

2.4 Materials and summary

The experiment was programmed using oTree (Chen et al., 2016) and conducted at the experimental laboratory of the Institute of Social and Economic Research (ISER) at Osaka University. 17 Sessions took place between

[5]Financial literacy scores across treatments were compared using an OLS regression model, regressing scores on treatment dummies with robust standard errors. An F-test was applied to compare the estimated dummy coefficients, with p-values used to report results.

October 2023 and June 2024, with each session dedicated to a single treatment. We recruited 311 Osaka University students from the ISER's ORSEE database (Greiner, 2015), and each participant attended only one session. Participants provided online consent by clicking an approval button prior to registering for the experiment. They received a participation fee of 500 JPY for completing the 90-minute session, with the potential to earn an additional reward of up to 1800 JPY based on their forecasting accuracy. Each session was facilitated by the same experimenter to maintain consistency.

In the main text, we report the average treatment effects without controlling for individual characteristics, as our results remained qualitatively consistent when controls were added. The detailed results with individual characteristic controls are available in the Online Appendix B.

2.5 Hypotheses

We hypothesized that participants were initially unaware of both their own forecasting accuracy and the performance of the advice source during Task 1. Thus, participants' WTP for advice in Task 1 would primarily reflect their perception of their forecasting abilities relative to those of the advice source. Bonaccio and Dalal (2006) reviews the literature on advice-giving and advice-taking, suggesting that that people generally prefer expert advice over non-expert advice, while Jussupow et al. (2020) summarizes findings suggesting that individuals are less likely to rely on algorithmic advice compared to expert recommendations in contexts like medical decision-making and stock price forecasting. Based on these findings, we propose the following hypothesis:

Hypothesis 1 *Participants' willingness to pay (WTP) for advice follows* *this order: (a) from experts > from students; (b) from experts > from algorithms.*

Upon completing Task 1, participants could assess their performance in both Parts 1 and 2 of the task. By comparing their initial performance (before accessing advice in Part 1) with their performance after adopting advice in Part 2, participants could evaluate any improvements attributable to the advice. If participants observed that the advice contributed positively to their performance in Task 1, we anticipate they would be more inclined to increase their WTP for advice in Task 2. Accordingly, we propose the following hypothesis:

Hypothesis 2 *Participants' WTP for advice will increase in Task 2 if they experience improved performance after adopting the advice in Task 1.*

3 Results

In this section, we present the results of the experiment. We begin by presenting characteristics of our participants. We then analyze participants' WTP for advice in Task 1, followed by an examination of WTP in Task 2 and changes between Tasks 1 and 2. We also explore the extent to which participants adopted the advice provided.

Table 2 summarizes the frequency of male participants, undergraduate grade level, science majors, stock trading experience, and financial literacy score. Characteristics of participants are well balanced across three treatments. Among the personal characteristics, only stock trading experience is statistically significantly different across the three treatments. In the Algorithm treatment, 20.755% of participants had stock trading experience,

Table 2: Summary of personal characteristics

Treatments	Algorithm	Expert	Student	p value
Male (%)	70.755	62.136	60.606	0.245
	(4.439)	(4.480)	(4.936)	
Undergraduate (%)	51.887	56.311	65.686	0.116
	(4.876)	(4.911)	(4.724)	
Science major (%)	74.528	67.961	68.627	0.507
	(4.252)	(4.620)	(4.617)	
Stock trading experience (%)	20.755	11.650	8.824	0.048
	(3.958)	(3.177)	(2.822)	
Financial literacy score (out of 12)	8.019	8.019	8.186	0.816
	(0.200)	(0.199)	(0.221)	
No. of participants	106	103	102	

Notes: a) Each personal characteristic was regressed on treatment dummies using an OLS regression model, with robust standard errors. The estimated dummy coefficients were compared using an F-test, with results presented as p-values.
b) Numbers in parentheses represent standard errors.
c) In the Student treatment, 3 participants refused to answer their gender, so the observation for Male (%) is based on 99 participants.

compared to 11.650% in the Expert treatment and 8.824% in the Student treatment.

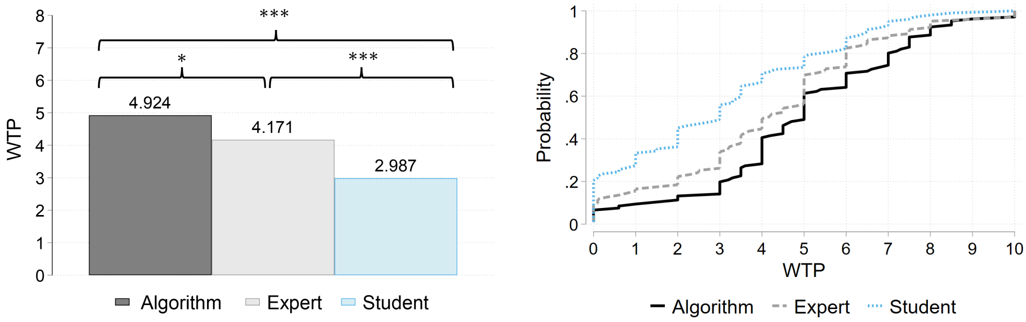
3.1 Willingness to Pay in Task 1

Figure 5 illustrates participants' WTP for advice in Task 1, where they were unaware of both their own performance and the performance of the advice source across the three treatments. Panel (a) presents the mean WTP for each treatment, while Panel (b) displays the cumulative distributions of WTPs.

The mean WTP to receive advice in Task 1, along with their standard deviations, were 4.924 (2.433) for the algorithm treatment, 4.171 (2.545) for the expert treatment, and 2.987 (2.508) for the student treatment. The mean WTP for algorithm-generated advice was marginally significantly higher than

Figure 5: WTP in Task 1

(a) Mean of WTP in Task 1 (b) CDF of WTP in Task 1



Notes: We regressed the WTP on three treatment dummies by OLS regression model with robust standard errors. Pair-wise comparisons of estimated dummy coefficients are done by F test with Bonferroni-adjusted p-values. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

that for expert advice ($p = 0.089$), while the mean WTP for expert advice was significantly higher than that for student advice ($p = 0.003$).[6] For this analysis, a two-tailed test was applied.

Panel (b) of Figure 5 reveals substantial variation in WTP within each treatment. Notably, each treatment includes participants with a WTP of 0. The proportion of participants with a WTP of 0 is highest in the student treatment (20.588%), more than twice that observed in the expert (8.738%) and algorithm (6.604%) treatments. The maximum WTP of 10 was observed for both the algorithm (3.774%) and expert (3.884%) treatments, whereas only 0.980% of participants in the student treatment reached this maximum.[7]

[6]The WTP were regressed on treatment dummies using an OLS regression model with robust standard errors. Pairwise comparisons of the estimated dummy coefficients were conducted using F-tests, with Bonferroni-adjusted p-values.

[7]We investigated whether personal characteristics correlate with WTP in Task 1 across different information sources, as this measure may reflect participants' self-assessed competence relative to the advice source. Online appendix Table B1 presents the results of the OLS regressions. For algorithmic advice, none of the observable personal characteristics exhibit a significant correlation with WTP in Task 1. In contrast, for expert advice, participants with stock trading experience show a significantly lower WTP (p = 0.016). For student advice, male participants exhibit a marginally significant decrease in WTP (p = 0.067). Other characteristics are not significantly correlated with WTP.

The WTP distribution for advice from students lies to the left of the distribution for expert advice, which in turn lies to the left of the distribution for algorithmic advice. We conducted a two-sample Kolmogorov–Smirnov test to assess distributional differences in WTP between treatments. The WTP distributions for algorithm and expert advice were statistically similar ($p = 0.123$), but the distributions significantly differed between algorithms and students ($p < 0.001$) and between experts and students ($p = 0.006$).

Therefore, we can state the following result.

Result 1 The WTP for advice follows the order: algorithm > experts > students. While Hypothesis 1-(a) is supported, Hypothesis 1-(b) is rejected.

Next, we examine WTP for advice in Task 2, where participants had the opportunity to observe both their own performance and any improvement in their performance after adopting the advice.

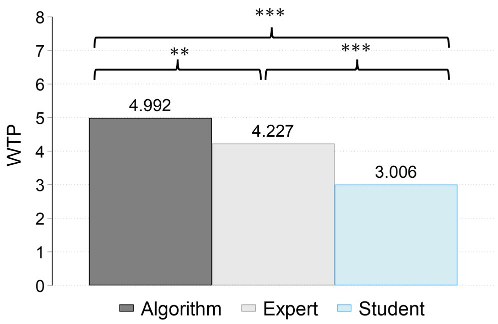
3.2 Effect of experiences on willingness to pay in Task 2

Figure 6 displays the WTP for advice across the three treatments in Task 2. Similar to Figure 5, Panel (a) presents the mean WTP, while Panel (b) illustrates the cumulative distributions. The ordering of both the mean values and the cumulative distributions of WTP in Task 2 closely resembles the patterns observed in Task 1.

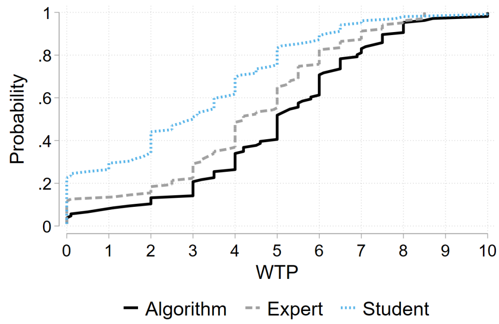
We are interested in, however, not only the aggregate result but also the results at the individual level. Specifically, we aim to understand whether participants' WTP values shifted between Tasks 1 and 2, and, if so, how these changes corresponded to their experiences in Task 1. To investigate this,

Figure 6: WTP in Task 2

(a) Mean of WTP in Task 2



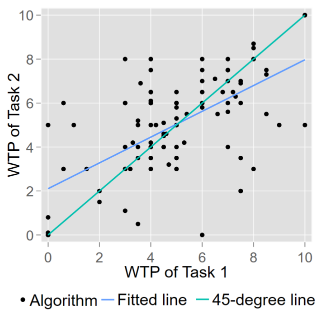
(b) CDF of WTP in Task 2



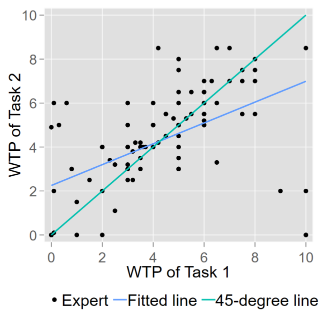
Notes: We regressed the WTP on three treatment dummies by OLS regression model with robust standard errors. Pair-wise comparisons of estimated dummy coefficients are done by F test with Bonferroni-adjusted p-values. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

Figure 7: Scatter plot of WTP in Task 1 vs. Task 2

(a) Algorithm



(b) Expert



(c) Student

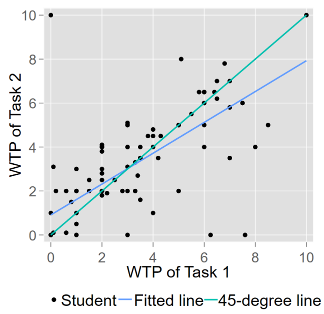


Figure 7 shows a scatter plot of WTP in Task 1 (x-axis) and Task 2 (y-axis). Each point represents a participant's WTPs in two tasks. Points located on the 45° line indicate participants whose WTP remained unchanged between the two tasks. Points above this line represent participants who increased their WTP in Task 2, whereas those below the line represent participants who decreased it.

Given the substantial variation in changes (or lack thereof) in WTP between the two tasks within each treatment, we now proceed with regression analyses. According to Hypothesis 2, the change in WTP between Tasks 1 and 2 should be positively correlated with the improvement in performance resulting from adopting advice in Task 1.

To test this hypothesis, we run a linear regression where the dependent variable $\triangle$WTP is the change in each participant's WTP between two tasks (Task 2 - Task 1). The main independent variable is the $\triangle$Mean performance in Task 1, which is the difference in the performance on the graphs when participants received the advice between two parts of Task 1 (Part 2 - Part 1). We also control for the mean price each participant paid by to receive the advice.

Additionally, in place of $\triangle$Mean performance in Task 1, we consider $\triangle$Net benefit in Task 1 as an alternative indepedent variable. This variable represents the increase in the payoff from Part 1 to Part 2, net of the price paid for advice.

Table 3 presents a summary of the regression results, with each participant serving as the unit of observation. We restricted the analysis to participants who received advice at least once. Models (1) to (3) consider $\triangle$Mean performance in Task 1 and the average cost of advice as independent variables, while models (4) to (6) consider $\triangle$Net benefit in Task 1 as the independent variable.

The estimated constant terms in Models (1) through (3) are positive and significant, indicating that, on average, participants increased their WTP from Task 1 to Task 2 across all treatments. The magnitude of the constant term is largest for the expert treatment, followed by the algorithm and student treatments. There is a statistically significant difference between the expert and student treatments ($p = 0.055$), but no significant difference

Table 3: Changes in WTP from Task 1 to Task 2

VARIABLES	(1) Algorithm $\triangle$WTP	(2) Expert $\triangle$WTP	(3) Student $\triangle$WTP	(4) Algorithm $\triangle$WTP	(5) Expert $\triangle$WTP	(6) Student $\triangle$WTP
$\triangle$Mean performance in Task 1	0.305***	0.219***	-0.003			
	(0.105)	(0.076)	(0.100)			
Average cost per advice paid in Task 1	-0.821***	-1.140***	-0.689***			
	(0.158)	(0.232)	(0.189)			
$\triangle$Net benefit in Task 1				0.415***	0.457***	0.244**
				(0.097)	(0.117)	(0.103)
Constant	2.184***	2.646***	1.175***	1.043***	0.517***	-0.066
	(0.444)	(0.540)	(0.332)	(0.248)	(0.150)	(0.187)
Obs.	97	84	73	97	84	73
R^2	0.330	0.448	0.263	0.251	0.253	0.076

Notes: a) In models 1-3, we regressed $\triangle$WTP on $\triangle$Mean performance from graphs with advice using an OLS regression model, controlling for the average cost per advice paid in Task 1, with robust standard errors. $\triangle$WTP is calculated as WTP in Task 2 minus WTP in Task 1. $\triangle$Mean performance is calculated as the mean performance from graphs with advice in Part 2 of Task 1 minus the mean performance from graphs with advice in Part 1 of Task 1. Performance ranges from 0 to 20, based on the points obtained in each question. In models 4-6, we regressed $\triangle$WTP on $\triangle$Net benefit from graphs with advice using an OLS regression model, with robust standard errors. $\triangle$Net benefit is calculated as the mean net benefit from graphs with advice in Part 2 of Task 1 minus the mean net benefit from graphs with advice in Part 1 of Task 1. The net benefit in each graph in Part 1 equals the rewards in each graph plus 10 points. The net benefit in each graph in Part 2 equals the rewards in each graph plus 10 points minus the price of advice. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.
b) The unit of observation is the participant who received at least one piece of advice in Task 1. The total number of observations is the number of participants who received at least one piece of advice in Task 1 in each treatment.
c) Numbers in parentheses represent standard errors.

between the expert and algorithm treatments ($p = 1.000$) or between the algorithm and student treatments ($p = 0.193$).[8]

We also observe negative and significant coefficients for the average cost per advice paid in Task 1. This indicates that, regardless of the advice source, a higher price paid for advice in Task 1 (reflecting a higher initial WTP) is associated with a smaller increase in WTP between Tasks 1 and 2. The absolute magnitude of this coefficient is largest in the expert treatment, followed by the algorithm treatment, and then the student treatment, although the differences across treatments are not statistically significant ($p = 0.301$).[9]

Regarding the effect of $\triangle$Mean performance in Task 1, the estimated coefficients are positive and significant for the algorithm and expert treatments but not significantly different from zero for the student treatment. While the reason for this non-significant coefficient in the student treatment is unclear, it suggests that both the initial WTP and the subsequent increase in WTP in response to improved performance after receiving advice may be influenced by the advice source.

It is worth noting, however, that the primary benefit of adopting advice is its effect on participants' payoffs. Thus, what matters is whether performance improvement due to advice is cost-effective. This aspect is examined in Models (4) through (6), where $\triangle$Net benefit in Task 1, which accounts for the cost of advice, serves as the independent variable.

In Models (4) through (6), the estimated constant terms are positive and significant for both the algorithm and expert treatments, with the constant term for the algorithm treatment being twice as large as that for the expert

[8] Comparisons of the constant terms across models were performed using an F-test, with Bonferroni-adjusted p-values provided for significance levels.

[9] Comparisons of the average cost per advice coefficient across models were conducted using an F-test, with results provided in p-values.

treatment. Conversely, the constant term for the student treatment is not significantly different from zero. The estimated coefficients for $\triangle$Net benefit in Task 1 are positive and significant across all three treatments.

These findings indicate that participants increased their WTP for advice from both the algorithm and expert sources when they observed positive change in performance as well as increase in Net benefit in Task 1. However, for the student treatment, participants only increased their WTP when the Net benefit increased in Task 1.

Result 2 Participants increased their WTP for algorithmic and expert advice after experiencing improved performance from adopting advice in Task 1, but not for student advice. Thus, Hypothesis 2 is partially supported.

Result 3 Participants increased their WTP for advice from algorithms, experts, and students after experiencing an improved net benefit from adopting advice in Task 1.

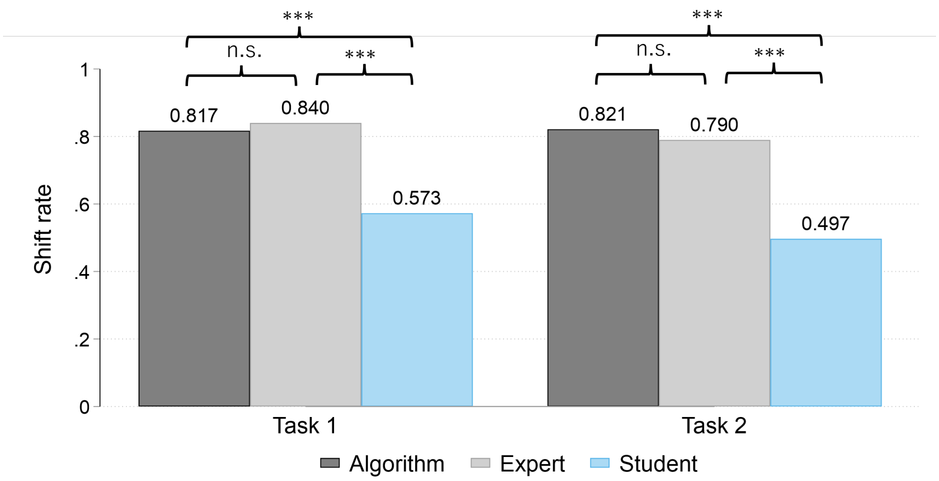
3.3 Shift rate

Let us now examine how the degree of advice utilization varies across different information sources. We quantify the degree of utilization using the *shift rate* which is a common measure in the literature on information utilization (Harvey and Fischer, 1997; Sniezek et al., 2004).[10] For each participant i and each graph g, the shift rate is defined as:

$$\text{shift rate}_g^i = \frac{\text{Final forecast}_g^i - \text{Initial forecast}_g^i}{\text{Advice}_g^i - \text{Initial forecast}_g^i}$$

[10]The shift rate in this paper is calculated in the same way as the weight of advice in the judge-advisor system paradigm, as used in previous studies (see, for example, the literature summarized in Bailey et al. (2023); Bonaccio and Dalal (2006)).

Figure 8: Mean shift rate



Notes: We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. Pair-wise comparisons of the estimated dummy coefficients are done by an F-test with Bonferroni-adjusted p-values. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, and n.s. means that the difference is not statistically significant at the 0.1 level.

where Initial forecast$_g^i$ and Final forecast$_g^i$ are the forecast submitted by participant i for graph g in Parts 1 and 2, respectively, and Advice$_g^i$ is the advice provided from the information source for graph g.

A shift rate of 0 indicates that participant i did not alter their forecast between Parts 1 and 2, regardless of the advice received. Conversely, a shift rate of 1 signifies that participant i fully incorporated the advice, aligning their Part 2 forecast exactly with the advice. Although shift rates can be less than 0 (indicating that the forecast in Part 2 moved in the opposite direction of the advice) or greater than 1 (indicating that the forecast in Part 2 exceeded the advice received), our focus is on cases where the shift rate falls between 0 and 1, inclusive.

Figure 8 displays the mean shift rates for the three treatments in Task 1 (left) and Task 2 (right). For both the algorithm and expert treatments, the mean shift rates are approximately 0.8 across both tasks, with no statistically significant difference between the two treatments. However, the mean shift rates for the student treatment are significantly lower, at 0.573 in Task 1 and 0.497 in Task 2.

This suggests that participants are not only willing to pay significantly less for advice from their peers (fellow students) but also tend to incorporate this advice less when they receive it. We now turn to investigating whether the shift rate is influenced by the price participants paid or their WTP, and whether these relationships vary by the source of advice. Table 4 presents the results of linear regressions where the shift rate is the dependent variable and the price paid for advice is an independent variable. In Models (4) through (6), WTP is also included as an independent variable.

Given the information source, the estimated coefficients of the price paid are not significantly different from zero in any of the models shown in Table 4. This holds true for WTP as well in Models (4) through (6). Furthermore, the shift rate is significantly lower in Task 2 for both the expert and student treatments, as indicated by the significant negative coefficients for the Task 2 dummy in Models (2), (3), (5), and (6). Thus, controlling for the information source, neither a higher price paid nor a greater WTP leads to increased utilization of a given piece of advice.

Additionally, we found no correlation between changes in shift rate and changes in WTP from Part 1 to Part 2 across all treatments ($p = 0.124$ for algorithm, $p = 0.151$ for experts, and $p = 0.348$ for students).[11]

[11] This analysis focused on cases where the shift rate is between 0 and 1. We regressed $\triangle$Shift rate on $\triangle$WTP using an OLS regression model with robust standard errors. Here, $\triangle$Shift rate is calculated as the mean shift rate for graphs with advice in Task 2 minus the mean shift rate for graphs with advice in Task 1, and $\triangle$WTP is calculated as WTP in Task 2 minus WTP in Task 1.

Table 4: Price of Advice and Shift Rate

VARIABLES	(1) Algorithm SHIFT	(2) Expert SHIFT	(3) Student SHIFT	(4) Algorithm SHIFT	(5) Expert SHIFT	(6) Student SHIFT
Price	0.001	0.002	0.003	-0.003	0.005	-0.003
	(0.005)	(0.007)	(0.008)	(0.004)	(0.005)	(0.010)
Task 2 dummy	0.004	-0.049*	-0.076**	0.005	-0.050*	-0.074**
	(0.021)	(0.026)	(0.034)	(0.020)	(0.026)	(0.033)
WTP				0.008	-0.006	0.009
				(0.009)	(0.012)	(0.016)
Constant	0.814***	0.834***	0.565***	0.775***	0.862***	0.535***
	(0.026)	(0.026)	(0.036)	(0.055)	(0.065)	(0.065)
Obs.	937	783	524	937	783	524
Clusters	101	93	80	101	93	80
R^2	0.000	0.008	0.014	0.003	0.010	0.016

Notes: a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between 0 and 1. We regressed shift rate on price of advice in each graph, using an OLS regression model, controlling for the Task 2 dummy and WTP, with robust standard errors clustered at participant level. Price indicates the random price of each graphs in each task. Task 2 dummy equals to 1 for Task 2 and 0 for Task 1. WTP indicates the submitted WTP in each task. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.
b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in Task 1 and Task 2 in each treatment.
c) Numbers in parentheses represent standard errors.

Result 4 Reliance on paid advice from the algorithm is similar to that from experts, but it is greater than reliance on advice from students.

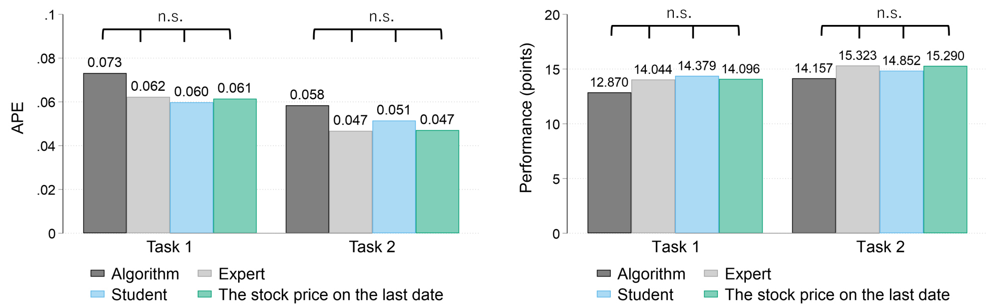
Result 5 Increasing the price of advice from algorithms, experts, and students does not, on average, influence the reliance on paid advice from these sources.

4 Discussion

So far, we have examined participants' WTP for advice from three different information sources and their utilization of this advice upon receipt. Contrary to our hypothesis, participants demonstrated a significantly higher

Figure 9: Performance of advice

(a) Absolute Percentage Error (APE) (b) Performance (points)



Notes: We regressed (a) the APE of advice and (b) the performance of advice in points on four dummy variable (three treatment dummies and last price dummy) using an OLS regression model with robust standard errors. We compared the estimated dummy coefficients using an F-test, with results illustrated by p-values. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, and n.s. denotes that the difference is not statistically significant at the 0.1 level.

WTP for advice from algorithm compared to expert advice. Additionally, we observed that participants' WTP increases in response to a rise in net benefits derived from utilizing the advice. However, we have not yet analyzed the resulting changes in performance or benefits from adopting the advice. Let us now turn to investigating the impact of advice on participants' performance.[12]

4.1 Performance of advice

Figure 9 illustrates the average performance of the three information sources, alongside the freely available information, i.e., the last price shown in the graph, as a benchmark. Panel (a) displays the mean absolute percentage error (APE), calculated over 10 forecasts in each task, for advice provided

[12]The analyses reported in this section are not pre-registered. We decided to include them based on feedback received during seminars and conferences.

by each source. The mean APE for a given source s is defined as:

$$\text{Mean APE}^s = \frac{1}{10}\sum_{g=1}^{10}\left|\frac{\text{Advice}_g^s - \text{True value}_g}{\text{True value}_g}\right|$$

where Advice_g^s is the advice provided by source s for graph g and True value_g denotes the actual target value for graph g. The mean APE is calculated by averaging over the 10 graphs in each task. As this measure represents the deviation of the advice from the true value, a higher APE indicates lower performance.

Panel (b) illustrates the average performance in terms of points participants could have earned if they had fully adopted the advice. Specifically, the mean performance points for source s are defined as follows, where $(\cdot)^+$ denotes $\max(\cdot, 0)$:

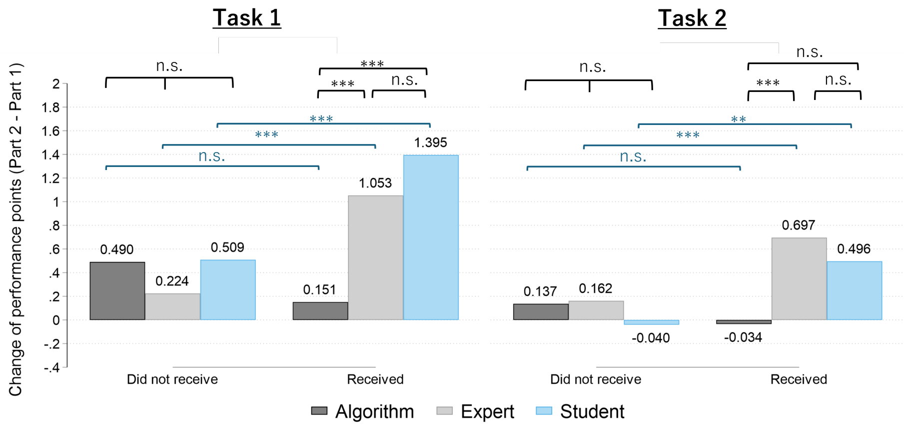
$$\text{Mean performance points}^s = \frac{1}{10}\sum_{g=1}^{10}\left(20 - \left|\frac{\text{Advice}_g^s - \text{True value}_g}{\text{True value}_g}\right| \times 100\right)^+$$

Thus, a higher mean performance point indicates better performance. This metric reflects the potential points participants could have gained by aligning their forecasts fully with the provided advice.

As evident in Figure 9, although the performance – as measured by both mean APE and mean performance points – does not significantly differ across the three information sources in either task, the algorithm actually performed the worst compared to the experts and students, whose performances were similar to each other across both tasks.[13] Even these two better-performing

[13]When we consider the performance of the advice actually received by participants, the average performance of the received advice, in terms of APE and performance points, is worse in the algorithm treatment compared to the expert treatment and the student treatment, and is similar between the expert treatment and the student treatment. Further details are provided in Online Appendix C1.

Figure 10: Change of performance (points)



Notes: We regressed the change in performance points (Part 2 - Part 1) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. Pairwise comparisons of the estimated dummy coefficients were performed using an F-test with Bonferroni-adjusted p-values. We also regressed the change in performance points (Part 2 - Part 1) on an advice-receiving dummy in each treatment using an OLS regression model with robust standard errors clustered at the participant level. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, while n.s. denotes that the difference is not statistically significant at the 0.1 level.

sources, however, are not any better than the last observed price, which is freely available.

Consequently, participants unknowingly expressed the highest willingness to pay for advice from the lowest-performing source. This tendency likely has adverse implications, both in terms of the accuracy of their final forecasts (submitted in Part 2) and in the payoffs they ultimately received.

Figure 10 displays the average change in performance, measured by points earned, from Part 1 to Part 2 of Task 1 (left panel) and Task 2 (right panel) for each of the three treatments.[14] Within each task panel, the average performance change for participants who did not receive advice is shown on the left, while those who received advice are shown on the right.

[14]The results regarding the change in APE are provided in Online Appendix C2.

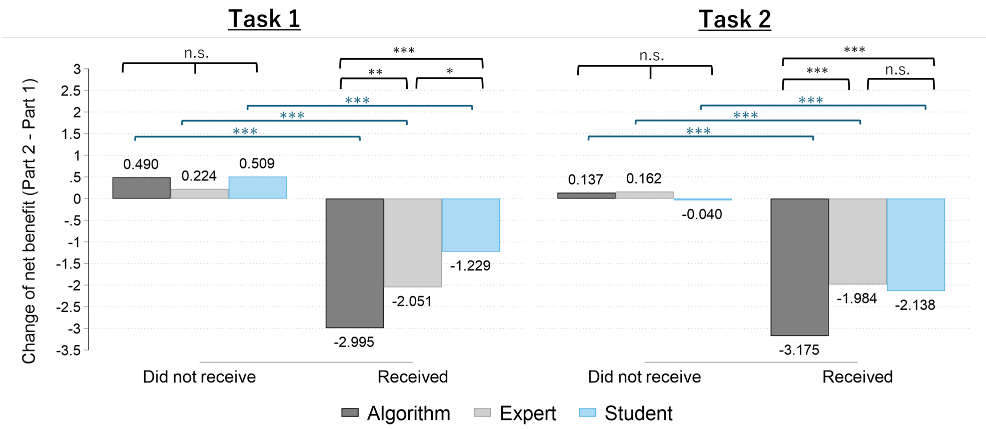
In both tasks, participants who did not receive advice exhibited slight performance improvements from Part 1 to Part 2 (except for those in the student treatment in Task 2). The mean performance increases were less than 1 point across all treatments, with no significant differences observed between them.

For participants who received advice from experts or students, the performance improvement was significantly greater than for those who did not receive advice. In Task 1, the mean performance increases were 1.053 points for the expert treatment and 1.395 points for the student treatment. However, for those who received advice from the algorithm, performance improved by an average of only 0.151 points, which was less than the 0.490-point improvement observed among those who did not receive advice. This difference was statistically insignificant. A similar pattern was observed in Task 2.

Figure 11 illustrates the average change in net benefit (payoff) from Part 1 to Part 2 in Task 1 (left) and Task 2 (right) for each of the three treatments. Accounting for the costs associated with receiving advice, these net benefit changes were, in fact, negative for participants who paid for advice. Recall that the average WTP for advice was approximately 5 for the algorithm, 4 for the expert, and 3 for the student sources. Reflecting this WTP along with the modest gain in points, the average decrease in payoff was largest for the algorithm treatment (-2.995), followed by the expert (-2.051) and student treatments (-1.229) in Task 1. In Task 2, advice from the algorithm continued to lead to the largest decrease (-3.175), while payoff reductions in the expert (-1.984) and student (-2.138) treatments also remained substantial.

Given the negative impact on payoffs from purchasing advice, it is somewhat perplexing that some participants increased their WTP in Task 2 com-

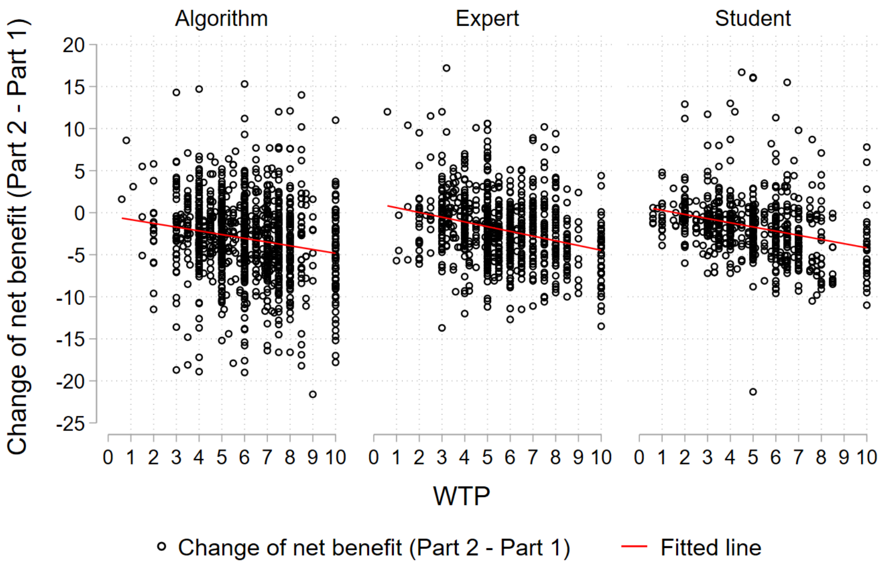
Figure 11: Change of net benefit (points)



Notes: We regressed the change of net benefit (Part 2 - Part 1) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. Pairwise comparisons of the estimated dummy coefficients were performed using an F-test with Bonferroni-adjusted p-values. We also regressed the change of net benefit (Part 2 - Part 1) on an advice-receiving dummy in each treatment using an OLS regression model with robust standard errors clustered at the participant level. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, and n.s. denotes that the difference is not statistically significant at the 0.1 level.

pared to Task 1. The proportion of participants who raised their WTP after observing the negative net benefit from advice in Task 1 was 30% for those advised by the algorithm, 21% for those advised by experts, and 16% for those advised by students. There is no significant difference between the algorithm and expert treatments ($p = 0.441$) or between the expert and student treatments ($p = 0.895$). However, there is a significant 5% difference between the algorithm and student treatments ($p = 0.044$).[15]

Figure 12: WTP vs. Change of net benefit in points (Part 2 - Part 1)



4.2 Maximum WTP to Increase Net Benefit

Figure 12 shows a scatter plot of WTP values and their corresponding changes in net benefits, measured in points (Part 2 - Part 1), across treatments, incorporating data from both Task 1 and Task 2. The red line represents the linear fit of the change in net benefit on WTP. To estimate this relationship, we regressed the change in net benefits on WTP using an OLS regression model with robust standard errors clustered at the participant level. From this model, we determined the WTP value at which the average change in net benefit equals zero.

The maximum WTP required to achieve an average increase in net benefit was 0 for the algorithm, 2.051 for experts, and 1.587 for students. This

[15] An irrational dummy variable was constructed, taking a value of 1 for participants who increased their WTP in Task 2 relative to Task 1, despite observing a negative net benefit from advice in Task 1, and 0 otherwise. We regressed the irrational dummy on three treatment dummies using a logit regression model with robust standard errors. Pairwise comparisons of the estimated dummy coefficients were performed using an F-test with Bonferroni-adjusted p-values.

result suggests that participants consistently overpaid for advice from all information sources, as their actual WTP exceeded the threshold required to increase net benefits.[16]

5 Conclusion

In this paper, we investigated (1) the willingness to pay (WTP) for stock price forecasts provided by an algorithm, financial experts, and peers, and (2) the extent to which participants utilized or relied on advice from these sources when making final decisions within a stock price forecasting experiment.

Our findings show that participants' WTP was highest for algorithmic advice, followed by expert advice, with the lowest WTP for peer advice. This pattern extended beyond WTP: participants also adopted algorithmic and expert advice to a significantly greater degree than peer advice, with no significant difference in reliance between the algorithm and expert sources. Our study reveals an inclination toward "algorithm appreciation" among participants.

This appreciation for algorithmic advice may reflect a shift in public perception, especially among university students, about AI's reliability and value. However, this preference came at a cost: the algorithm did not outperform other sources and, in some cases, performed worse. Consequently, participants' greater WTP for and reliance on algorithmic advice reduced their net payoffs, as the value added by the algorithm's guidance did not justify its cost.

Some participants did not fully recognize that they were overpaying for

[16]The results on the relationship between WTP and net benefits in Part 2 are provided in Online Appendix C3, while the relationship between WTP and changes in performance is discussed in Online Appendix C4.

advice, irrespective of the source. These findings underscore the need for policies and tools that enhance individuals' ability to assess the performance of information sources, particularly algorithms.

Given participants' tendency to overvalue algorithmic forecasts despite their mixed performance, regulatory bodies could establish standards for transparent disclosures about AI performance, particularly in finance and other high-stakes areas. Clear, accessible information on the accuracy and limitations of AI-driven predictions would help mitigate over-reliance on AI, encouraging better-informed decision-making. Additionally, policies that foster AI literacy –through educational programs and certification standards in both financial literacy and AI competency –could help consumers and professionals critically evaluate when to rely on algorithmic or human expertise. Such initiatives would help protect consumer interests, improve decision quality, and support a balanced integration of AI across various sectors.

References

Bailey, P. E., T. Leon, N. C. Ebner, A. A. Moustafa, and G. Weidemann (2023): "A meta-analysis of the weight of advice in decision-making," *Current Psychology*, 42, 24516–24541.

Bao, T., B. Corgnet, N. Hanaki, K. Okada, Y. E. Riyanto, and J. Zhu (2022): "Financial Forecasting in the Lab and the Field: Qualified Professionals vs. Smart Students," Iser dp 1156, Institute of Social and Economic Research, Osaka University.

Bao, T., B. Corgnet, N. Hanaki, Y. E. Riyanto, and J. Zhu (2023):

"Predicting the unpredictable: New experimental evidence on forecasting random walks," *Journal of Economic Dynamics and Control*, 146, 104571.

Becker, G. M., M. H. DeGroot, and J. Maschak (1964): "Measuring utility by a single-response sequential method," *Behavioral Science*, 9, 226–232.

Bonaccio, S. and R. S. Dalal (2006): "Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences," *Organizational Behavior and Human Decision Processes*, 101, 127–151.

Burton, J. W., M.-K. Stein, and T. B. Jensen (2020): "A systematic review of algorithm aversion in augmented decision making," *Journal of Behavioral Decision Making*, 33, 220–239.

Capraro, C., A. Lentsch, D. Acemoglu, S. Akgun, A. Akhmedova, E. Bilancini, J.-F. Bonnefon, P. B. nas Garza, L. Butera, K. M. Douglas, J. A. Everett, G. Gigerenzer, C. Greenhow, D. A. Hashimoto, J. Holt-Lunstad, J. Jetten, S. Johnson, W. H. Kunz, C. Longoni, P. Lunn, S. Natale, S. Paluch, I. Rahwan, N. Selwyn, V. Singh, S. Suri, J. Sutcliffe, J. Tomlinson, S. "van der Linden"', P. A. M. V. Lange"', F. Wall, J. J. V. Bavel;;, and R. Viale (2024): "The impact of generative artificial intelligence on socioeconomic inequalities and policy making," *PNAS nexus*, 3, 191.

Cason, T. N. and C. R. Plott (2014): "Misconceptions and Game Form

Recognition: Challenges to Theories of Revealed Preference and Framing," *Journal of Political Economy*, 122, 1235–1270.

Castelo, N., M. W. Bos, and D. R. Lehmann (2019): "Task-dependent algorithm aversion," *Journal of Marketing Research*, 56, 809–825.

Chacon, A., E. E. Kausel, and T. Reyes (2022): "A longitudinal approach for understanding algorithm use," *Journal of Behavioral Decision Making*, 35, e2275.

Chen, D. L., M. Schonger, and C. Wickens (2016): "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

D'Acunto, F., N. Prabhala, and A. G. Rossi (2019): "The promises and pitfalls of robo-advising," *The Review of Financial Studies*, 32, 1983–2020.

Dietvorst, B. J., J. P. Simmons, and C. Massey (2015): "Algorithm aversion: people erroneously avoid algorithms after seeing them err." *Journal of Experimental Psychology: General*, 144, 114.

——— (2018): "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them," *Management Science*, 64, 1155–1170.

Dimmock, S. G., W. C. Gerken, and N. P. Graham (2018): "Is fraud contagious? Coworker influence on misconduct by financial advisors," *The Journal of Finance*, 73, 1417–1450.

Egan, M., G. Matvos, and A. Seru (2019): “The market for financial adviser misconduct,” *Journal of Political Economy*, 127, 233–295.

Fernandes, D., J. G. L. Jr., and R. G. Netemeyer (2014): “Financial literacy, financial education, and downstream financial behaviors,” *Management Science*, 60, 1861–1883.

Filiz, I., J. R. Judek, M. Lorenz, and M. Spiwoks (2021): “Reducing algorithm aversion through experience,” *Journal of Behavioral and Experimental Finance*, 31, 100524.

Foerster, S., J. T. Linnainmaa, B. T. Melzer, and A. Previtero (2017): “Retail financial advice: does one size fit all?” *The Journal of Finance*, 72, 1441–1482.

Freer, M., D. Friedman, and S. Weidenholzer (2023): “Motives for delegating financial decisions,” *arXiv preprint arXiv:2309.03419*.

Germann, M. and C. Merkle (2023): “Algorithm aversion in delegated investing,” *Journal of Business Economics*, 93, 1691–1727.

Gill, A., R. M. Gillenkirch, J. Ortner, and L. Velthuis (2024): “Dynamics of Reliance on Algorithmic Advice,” *Journal of Behavioral Decision Making*, 37, e2414.

Gino, F. (2008): “Do we listen to advice just because we paid for it? The impact of advice cost on its use,” *Organizational Behavior and Human Decision Processes*, 107, 234–245.

Greiner, B. (2015): “Subject pool recruitment procedures: Organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

Greiner, B., P. Grünwald, T. Lindner, G. Lintner, and M. Wiernsperger (2024): "Incentives, framing, and reliance on algorithmic advice: An experimental study," Department of Strategy and Innovation Working Paper 01/2024, Vienna University of Economics and Business.

Harvey, N. and I. Fischer (1997): "Taking advice: Accepting help, improving judgment, and sharing responsibility," *Organizational behavior and human decision processes*, 70, 117–133.

Himmelstein, M. and D. V. Budescu (2023): "Preference for human or algorithmic forecasting advice does not predict if and how it is used," *Journal of Behavioral Decision Making*, 36, e2285.

Hoechle, D., S. Ruenzi, N. Schaub, and M. Schmid (2018): "Financial advice and bank profits," *The Review of Financial Studies*, 31, 4447–4492.

Holzmeister, F., M. Holmén, M. Kirchler, M. Stefan, and E. Wengström (2023): "Delegation decisions in finance," *Management Science*, 69, 4828–4844.

Jussupow, E., I. Benbasat, and A. Heinzl (2020): "Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion," in *In Proceedings of the 28th European Conference on Information Systems (ECIS).*

Korinek, A. (2024): "Economic policy challenges for the age of AI," Workig Paper 32980, National Bureau of Economic Research.

Linnainmaa, J. T., B. T. Melzer, and A. Previtero (2021): "The misguided beliefs of financial advisors," *The Journal of Finance*, 76, 587–621.

Logg, J. M., J. A. Minson, and D. A. Moore (2019): "Algorithm appreciation: People prefer algorithmic to human judgment," *Organizational Behavior and Human Decision Processes*, 151, 90–103.

Mahmud, H., A. N. Islam, S. I. Ahmed, and K. Smolander (2022): "What influences algorithmic decision-making? A systematic literature review on algorithm aversion," *Technological Forecasting and Social Change*, 175, 121390.

Noy, S. and W. Zhang (2023): "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 381, 187–192.

OECD (2019): "Artificial Intelligence in Society," *OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en.*

——— (2024): "Regulatory approaches to Artificial Intelligence in finance," *OECD Publishing, Paris, https://doi.org/https://doi.org/10.1787/f1498c02-en.*

Prahl, A. and L. Van Swol (2017): "Understanding algorithm aversion: When is advice from automation discounted?" *Journal of Forecasting*, 36, 691–702.

Sirri, E. R. and P. Tufano (1998): "Costly search and mutual fund flows," *The Journal of Finance*, 53, 1589–1622.

Sniezek, J. A., G. E. Schrah, and R. S. Dalal (2004): "Improving judgement with prepaid expert advice," *Journal of Behavioral Decision Making*, 17, 173–190.

Sutherland, S. C., C. Harteveld, and M. E. Young (2016): "Effects of the advisor and environment on requesting and complying with automated advice," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 6, 1–36.

Tse, T. T. K., N. Hanaki, and B. Mao (2024): "Beware the performance of an algorithm before relying on it: Evidence from a stock price forecasting experiment," *Journal of Economic Psychology*, 102, 102727.

Online Appendix to "Valuing Algorithms Over Experts: Evidence from a Stock Price Forecasting Experiment"

Nobuyuki Hanaki* Bolin Mao † Tiffany Tsz Kwan Tse ‡ Wenxin Zhou §

Contents

*Corresponding author: Institute of Social and Economic Research, Osaka University, Japan and University of Limassol, Cyprus. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

†Graduate School of Economics, Hitotsubashi University, Japan. E-mail: meta.bolin.mao@gmail.com

‡Institute of Social and Economic Research, Osaka University, Japan. E-mail: tiffany.econ@gmail.com

§A non-academic institution. E-mail: haruka_ivory@outlook.com

A Results for Figures

Table A1: Predicted WTP in Task 1 and Task 2 (for Figure 5 and 6)

VARIABLES	(1) WTP Task 1	(2) WTP Task 2
Baseline = AI treatment dummy		
Expert treatment dummy	-0.753**	-0.765**
	(0.345)	(0.317)
Student treatment dummy	-1.937***	-1.987***
	(0.343)	(0.329)
Constant	4.924***	4.992***
	(0.236)	(0.222)
Observations	311	311
R-squared	0.093	0.110
Treatments	Predicted WTP Task 1	Predicted WTP Task 2
AI	4.924	4.992
	(0.236)	(0.222)
Experts	4.171	4.227
	(0.251)	(0.226)
Students	2.987	3.006
	(0.248)	(0.243)
	Prob > F	Prob > F
AI = Experts = Students	0.000	0.000
AI = Expert #	0.089	0.049
AI = Student #	0.000	0.000
Expert = Student #	0.003	0.001

Notes: (a) We regressed the WTP on three treatment dummies by OLS regression model with robust standard errors. (b) The unit of observation is the number of participants. The total number of observations is the number of participants in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table A2: Predicted shift rate in Task 1 and Task 2 (for Figure 8)

VARIABLES	(1) Shift rate Task 1	(2) Shift rate Task 2
Baseline = AI treatment dummy		
Expert treatment dummy	0.023	-0.032
	(0.031)	(0.033)
Student treatment dummy	-0.244***	-0.325***
	(0.038)	(0.033)
Constant	0.817***	0.821***
	(0.021)	(0.021)
Observations	1,118	1,126
Clusters	251	259
R-squared	0.123	0.178
Treatments	Predicted shift rate Task 1	Predicted shift rate Task 2
AI	0.817	0.821
	(0.021)	(0.021)
Experts	0.840	0.790
	(0.023)	(0.025)
Students	0.573	0.497
	(0.031)	(0.026)
	Prob > F	Prob > F
AI = Experts = Students	0.000	0.000
AI = Expert #	1.000	1.000
AI = Student #	0.000	0.000
Expert = Student #	0.000	0.000

Notes: (a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between 0 and 1. We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table A3: Performance of advice in Task 1 and Task 2 (for Figure 9)

VARIABLES	(1) APE Task 1	(2) APE Task 2	(3) Points Task 1	(4) Points Task 2
Baseline = AI treatment dummy				
Expert treatment dummy	-0.011	-0.012	1.174	1.166
	(0.029)	(0.017)	(2.632)	(1.695)
Student treatment dummy	-0.013	-0.007	1.509	0.695
	(0.029)	(0.016)	(2.597)	(1.576)
Constant	0.073***	0.058***	12.870***	14.157***
	(0.021)	(0.013)	(1.975)	(1.268)
Observations	30	30	30	30
R-squared	0.009	0.020	0.014	0.020
Treatments	Predicted APE Task 1	Predicted APE Task 2	Predicted Points Task 1	Predicted Points Task 2
AI	0.073	0.058	12.870	14.157
	(0.021)	(0.013)	(1.975)	(1.268)
Experts	0.062	0.047	14.044	15.323
	(0.020)	(0.011)	(1.740)	(1.124)
Students	0.060	0.051	14.379	14.852
	(0.020)	(0.009)	(1.687)	(0.936)
	Prob > F	Prob > F	Prob > F	Prob > F
AI = Experts = Students	0.891	0.790	0.838	0.790

Notes: (a) We regressed the APE of advice / the performance of advice in points on three treatment dummies using an OLS regression model with robust standard errors. (b) The unit of observation is the prediction of a graph from each information source. The total number of observations is the total number of graphs in all treatments. (c) The robust standard errors are in parentheses.

Table A4: Change of performance (points) (for Figure 10)

VARIABLES	Advice received	(1) ΔPerf. Task 1 No	(2) ΔPerf. Task 1 Yes	(3) ΔPerf. Task 2 No	(4) ΔPerf. Task 2 Yes
Baseline = AI treatment dummy					
Expert treatment dummy		-0.266	0.901***	0.025	0.731***
		(0.243)	(0.279)	(0.147)	(0.228)
Student treatment dummy		0.019	1.244***	-0.177	0.530*
		(0.193)	(0.267)	(0.144)	(0.274)
Constant		0.490***	0.151	0.137	-0.034
		(0.157)	(0.206)	(0.114)	(0.174)
Observations		1,867	1,243	1,854	1,256
Clusters		298	254	305	261
R-squared		0.001	0.015	0.001	0.008
Treatments		Predicted ΔPerf. Task 1	Predicted ΔPerf. Task 2	Predicted ΔPerf. Task 1	Predicted ΔPerf. Task 2
AI		0.490	0.151	0.137	-0.034
		(0.157)	(0.206)	(0.114)	(0.174)
Experts		0.224	1.053	0.162	0.697
		(0.186)	(0.189)	(0.093)	(0.147)
Students		0.509	1.395	-0.040	0.496
		(0.113)	(0.171)	(0.088)	(0.211)
		Prob > F	Prob > F	Prob > F	Prob > F
AI = Experts = Students		0.406	0.000	0.239	0.006
AI = Expert #			0.004		0.005
AI = Student #			0.000		0.162
Expert = Student #			0.540		1.000

Notes: (a) We regressed the change in performance points (Part 2 - Part 1) (ΔPerf.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table A5: Change in performance (points) when comparing with and without receiving advice (for Figure 10)

VARIABLES	(1) ΔPerf. Task 1 Algorithm	(2) ΔPerf. Task 1 Expert	(3) ΔPerf. Task 1 Student	(4) ΔPerf. Task 2 Algorithm	(5) ΔPerf. Task 2 Expert	(6) ΔPerf. Task 2 Student
Advice received dummy	-0.339	0.828***	0.886***	-0.171	0.535***	0.536**
	(0.249)	(0.259)	(0.195)	(0.217)	(0.161)	(0.232)
Constant	0.490***	0.224	0.509***	0.137	0.162*	-0.040
	(0.157)	(0.187)	(0.113)	(0.114)	(0.093)	(0.088)
Observations	1,060	1,030	1,020	1,060	1,030	1,020
Clusters	106	103	102	106	103	102
R-squared	0.002	0.011	0.015	0.001	0.009	0.008

Notes: (a) We regressed the change in performance points (Part 2 - Part 1) (ΔPerf.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecast with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses.

Table A6: Change of net benefit (points) (for Figure 11)

VARIABLES	Advice received	(1) ΔNB. Task 1 No	(2) ΔNB. Task 1 Yes	(3) ΔNB. Task 2 No	(4) ΔNB. Task 2 Yes
Baseline = AI treatment dummy					
Expert treatment dummy		-0.266	0.944***	0.025	1.192***
		(0.243)	(0.355)	(0.147)	(0.286)
Student treatment dummy		0.019	1.765***	-0.177	1.037***
		(0.193)	(0.345)	(0.144)	(0.349)
Constant		0.490***	-2.995***	0.137	-3.175***
		(0.157)	(0.232)	(0.114)	(0.213)
Observations		1,867	1,243	1,854	1,256
Clusters		298	254	305	261
R-squared		0.001	0.023	0.001	0.019
Treatments		Predicted ΔNB. Task 1	Predicted ΔNB. Task 2	Predicted ΔNB. Task 1	Predicted ΔNB. Task 2
AI		0.490	-2.995	0.137	-3.175
		(0.157)	(0.232)	(0.114)	(0.213)
Experts		0.224	-2.051	0.162	-1.984
		(0.186)	(0.269)	(0.093)	(0.191)
Students		0.509	-1.229	-0.040	-2.138
		(0.113)	(0.255)	(0.088)	(0.277)
		Prob > F	Prob > F	Prob > F	Prob > F
AI = Experts = Students		0.406	0.000	0.239	0.000
AI = Expert #			0.025		0.000
AI = Student #			0.000		0.0097
Expert = Student #			0.082		1.000

Notes: (a) We regressed the change in net benefit (Part 2 - Part 1) (ΔNB.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table A7: Change in net benefit (points) when comparing with and without receiving advice (for Figure 11)

VARIABLES	(1) ΔNB. Task 1 Algorithm	(2) ΔNB. Task 1 Expert	(3) ΔNB. Task 1 Student	(4) ΔNB. Task 2 Algorithm	(5) ΔNB. Task 2 Expert	(6) ΔNB. Task 2 Student
Advice received dummy	-3.485***	-2.275***	-1.738***	-3.312***	-2.146***	-2.098***
	(0.276)	(0.329)	(0.276)	(0.250)	(0.194)	(0.292)
Constant	0.490***	0.224	0.509***	0.137	0.162*	-0.040
	(0.157)	(0.187)	(0.113)	(0.114)	(0.093)	(0.088)
Observations	1,060	1,030	1,020	1,060	1,030	1,020
Clusters	106	103	102	106	103	102
R-squared	0.135	0.073	0.051	0.177	0.111	0.104

Notes: (a) We regressed the change in net benefit (Part 2 - Part 1) (ΔNB.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecast with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses.

Table A8: Summary of variables

VARIABLES	Algorithm Task 1	Algorithm Task 2	Expert Task 1	Expert Task 2	Student Task 1	Student Task 2
Average WTP	4.924	4.992	4.171	4.227	2.987	3.006
	(0.236)	(0.222)	(0.251)	(0.226)	(0.248)	(0.243)
Frequency of WTP=0 (%)	6.604	3.774	8.738	11.650	20.588	22.549
	(2.424)	(1.860)	(2.796)	(3.177)	(4.023)	(4.158)
Frequency of WTP=10 (%)	3.774	2.830	3.884	0	0.980	1.961
	(1.860)	(1.618)	(1.913)		(0.980)	(1.380)
Average price of paid advice	3.146	3.141	3.104	2.680	2.625	2.634
	(0.087)	(0.091)	(0.106)	(0.088)	(0.120)	(0.105)
Frequency of receiving advice (%) including WTP=0	50.471	47.547	39.417	42.136	29.608	31.176
	(1.536)	(1.535)	(1.523)	(1.539)	(1.430)	(1.451)
Frequency of receiving advice (%) when WTP$\neq$0	54.040	49.412	43.191	47.692	37.284	40.253
	(1.585)	(1.566)	(1.616)	(1.657)	(1.700)	(1.746)
Performance of advice (APE)	0.073	0.058	0.062	0.047	0.060	0.051
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Performance of advice (points)	12.870	14.157	14.044	15.323	14.379	14.852
	(0.182)	(0.117)	(0.160)	(0.104)	(0.155)	(0.086)
Performance (points) in Part 1	12.844	14.273	12.934	14.397	12.839	14.354
	(0.186)	(0.133)	(0.191)	(0.137)	(0.196)	(0.134)
Performance (points) in Part 2	13.163	14.328	13.485	14.784	13.610	14.481
	(0.182)	(0.124)	(0.177)	(0.121)	(0.182)	(0.126)
Improvement of performance (points) with advice	0.151	-0.034	1.053	0.697	1.395	0.496
	(0.213)	(0.180)	(0.190)	(0.151)	(0.203)	(0.178)
Improvement of performance (points) without advice	0.490	0.137	0.224	0.162	0.509	-0.040
	(0.152)	(0.102)	(0.149)	(0.098)	(0.118)	(0.093)
Improvement of net benefit (points) with advice	-2.995	-3.175	-2.051	-1.984	-1.229	-2.138
	(0.223)	(0.201)	(0.213)	(0.176)	(0.230)	(0.201)
Improvement of net benefit (points) without advice	0.490	0.137	0.224	0.162	0.509	-0.040
	(0.152)	(0.102)	(0.149)	(0.098)	(0.118)	(0.093)

Notes: Standard errors are in parentheses.

B Analyses of experimental results conditional on personal characteristics

We investigated whether personal characteristics correlate with WTP in Task 1, as this measure may reflect participants' perceptions of their own skills relative to the advice source. Table B1 presents these results. Models (1) through (3) use WTP in Task 1 as the dependent variable.

Table B1: Personal characteristics and WTP in Task 1

VARIABLES	(1) Algorithm WTP_{Task1}	(2) Expert WTP_{Task1}	(3) Student WTP_{Task1}
Male dummy	-0.315	-0.0134	-1.018*
	(0.585)	(0.533)	(0.549)
Undergraduate student dummy	0.0428	0.934	-0.251
	(0.478)	(0.564)	(0.601)
Science major dummy	0.461	-0.0461	0.0103
	(0.637)	(0.630)	(0.586)
Stock trading experience dummy	0.362	-1.753**	0.613
	(0.669)	(0.715)	(1.188)
Financial literacy score	-0.0691	0.129	0.0599
	(0.122)	(0.148)	(0.125)
Constant	5.260***	2.851*	3.185**
	(1.158)	(1.461)	(1.278)
Obs.	106	103	99
R^2	0.012	0.090	0.045

Notes: a) We regressed WTP in Task 1 on personal characteristics using an OLS regression model, with robust standard errors. The male dummy equals 1 for male participants and 0 otherwise. The undergraduate student dummy equals 1 for undergraduate students and 0 otherwise. The science major dummy equals 1 for science majors and 0 otherwise. The stock trading experience dummy equals 1 for participants with stock trading experience and 0 otherwise. The financial literacy score ranges from 0 to 12. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.
b) The unit of observation is the participant. The total number of observations corresponds to the number of participants in each treatment.
c) Numbers in parentheses represent standard errors.

For algorithmic advice, none of the observable personal characteristics show a significant correlation with WTP in Task 1. In contrast, for expert advice, participants with stock trading experience exhibit a lower WTP, with a statistically significant result ($p = 0.016$). For student advice, male participants demonstrate a marginally significant decrease in WTP ($p = 0.067$).

Overall, most personal characteristics do not significantly correlate with WTP for advice from any source. While this study focused on a limited set of personal characteristics, future research could expand this exploration to additional attributes to assess their potential impact on WTP for advice from various information sources.

Table B2: Predicted WTP in Task 1 and Task 2 conditional on personal characteristics

VARIABLES	(1) WTP Task 1	(2) WTP Task 2
Baseline = AI treatment dummy		
Expert treatment dummy	-0.820**	-0.877***
	(0.353)	(0.316)
Student treatment dummy	-2.075***	-2.078***
	(0.360)	(0.330)
Male dummy	-0.388	-0.596**
	(0.318)	(0.296)
Undergraduate student dummy	0.192	-0.224
	(0.312)	(0.288)
Science major dummy	-0.041	-0.563*
	(0.351)	(0.320)
Stock trading experience dummy	-0.245	-0.370
	(0.451)	(0.409)
Financial literacy score	0.037	-0.046
	(0.073)	(0.066)
Constant	4.884***	6.397***
	(0.723)	(0.612)
Observations	308	308
R-squared	0.104	0.144
Treatments	Predicted WTP Task 1	Predicted WTP Task 2
AI	4.980	5.061
	(0.242)	(0.221)
Experts	4.160	4.184
	(0.249)	(0.224)
Students	2.905	2.983
	(0.256)	(0.244)
	Prob > F	Prob > F
AI = Experts = Students	0.000	0.000
AI = Expert #	0.063	0.018
AI = Student #	0.000	0.000
Expert = Student #	0.001	0.001

Notes: (a) We regressed the WTP on three treatment dummies by OLS regression model with robust standard errors, controlling personal characteristics. (b) The unit of observation is the number of participants. The total number of observations is the number of participants in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table B3: Predicted shift rate in Task 1 and Task 2 conditional on personal characteristics

VARIABLES	(1) Shift rate Task 1	(2) Shift rate Task 2
Baseline = AI treatment dummy		
Expert treatment dummy	0.038	-0.029
	(0.030)	(0.033)
Student treatment dummy	-0.235***	-0.329***
	(0.039)	(0.035)
Male dummy	-0.041	-0.025
	(0.031)	(0.031)
Undergraduate student dummy	0.026	0.035
	(0.031)	(0.029)
Science major dummy	0.048	0.005
	(0.036)	(0.031)
Stock trading experience dummy	0.086**	0.051
	(0.039)	(0.037)
Financial literacy score	-0.012	0.004
	(0.008)	(0.007)
Constant	0.869***	0.772***
	(0.078)	(0.068)
Observations	1,107	1,119
Clusters	249	257
R-squared	0.138	0.186
Treatments	Predicted shift rate Task 1	Predicted shift rate Task 2
AI	0.810	0.821
	(0.021)	(0.022)
Experts	0.848	0.792
	(0.021)	(0.025)
Students	0.575	0.491
	(0.032)	(0.027)
	Prob > F	Prob > F
AI = Experts = Students	0.000	0.000
AI = Expert #	0.648	1.000
AI = Student #	0.000	0.000
Expert = Student #	0.000	0.000

Notes: (a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between 0 and 1. We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. (b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table B4: Change of performance (points) conditional on personal characteristics

VARIABLES	Advice received	(1) ΔPerf. Task 1 No	(2) ΔPerf. Task 1 Yes	(3) ΔPerf. Task 2 No	(4) ΔPerf. Task 2 Yes
Baseline = AI treatment dummy					
Expert treatment dummy		-0.266	0.821***	0.029	0.752***
		(0.253)	(0.271)	(0.148)	(0.232)
Student treatment dummy		0.010	1.148***	-0.133	0.530*
		(0.203)	(0.277)	(0.140)	(0.277)
Male dummy		-0.004	-0.084	0.096	-0.285
		(0.204)	(0.252)	(0.120)	(0.234)
Undergraduate student dummy		-0.032	0.209	-0.111	-0.198
		(0.186)	(0.235)	(0.113)	(0.238)
Science major dummy		-0.158	-0.411	-0.125	-0.221
		(0.232)	(0.299)	(0.139)	(0.243)
Stock trading experience dummy		0.059	-0.085	0.059	0.292
		(0.298)	(0.373)	(0.178)	(0.321)
Financial literacy score		0.021	-0.001	-0.030	0.015
		(0.042)	(0.053)	(0.031)	(0.048)
Constant		0.447	0.439	0.454	0.227
		(0.475)	(0.552)	(0.315)	(0.468)
Observations		1,851	1,229	1,832	1,248
Clusters		296	252	302	259
R-squared		0.002	0.018	0.002	0.013
Treatments		Predicted ΔPerf. Task 1	Predicted ΔPerf. Task 2	Predicted ΔPerf. Task 1	Predicted ΔPerf. Task 2
AI		0.489	0.195	0.123	-0.041
		(0.161)	(0.206)	(0.113)	(0.177)
Experts		0.223	1.017	0.152	0.711
		(0.186)	(0.184)	(0.093)	(0.148)
Students		0.498	1.343	-0.010	0.488
		(0.118)	(0.184)	(0.086)	(0.211)
		Prob > F	Prob > F	Prob > F	Prob > F
AI = Experts = Students		0.421	0.000	0.383	0.005
AI = Expert #			0.008		0.004
AI = Student #			0.000		0.171
Expert = Student #			0.642		1.000

Notes: (a) We regressed the change in performance points (Part 2 - Part 1) (ΔPerf.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table B5: Change in performance (points) when comparing with and without receiving advice conditional on personal characteristics

VARIABLES	(1) ΔPerf. Task 1 Algorithm	(2) ΔPerf. Task 1 Expert	(3) ΔPerf. Task 1 Student	(4) ΔPerf. Task 2 Algorithm	(5) ΔPerf. Task 2 Expert	(6) ΔPerf. Task 2 Student
Advice received dummy	-0.352	0.753***	0.836***	-0.209	0.562***	0.476**
	(0.256)	(0.245)	(0.194)	(0.220)	(0.158)	(0.233)
Male dummy	-0.092	0.196	-0.302	0.025	-0.068	-0.129
	(0.307)	(0.304)	(0.225)	(0.224)	(0.202)	(0.197)
Undergraduate student dummy	0.444*	0.283	-0.658***	-0.386*	0.117	-0.180
	(0.263)	(0.265)	(0.226)	(0.205)	(0.191)	(0.222)
Science major dummy	-0.272	-0.597	0.279	-0.393*	-0.051	-0.296
	(0.376)	(0.376)	(0.216)	(0.228)	(0.229)	(0.205)
Stock trading experience dummy	0.341	-0.276	-0.187	-0.155	0.719*	-0.165
	(0.340)	(0.482)	(0.352)	(0.241)	(0.364)	(0.298)
Financial literacy score	-0.039	0.010	0.031	-0.046	-0.044	0.052
	(0.066)	(0.066)	(0.049)	(0.046)	(0.053)	(0.046)
Constant	0.775	0.326	0.706	1.033**	0.433	-0.021
	(0.728)	(0.673)	(0.486)	(0.442)	(0.510)	(0.497)
Observations	1,060	1,030	990	1,060	1,030	990
Clusters	106	103	99	106	103	99
R-squared	0.007	0.021	0.027	0.006	0.016	0.014

Notes: (a) We regressed the change in performance points (Part 2 - Part 1) (ΔPerf.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecast with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses.

Table B6: Change of net benefit (points) conditional on personal characteristics

VARIABLES	Advice received	(1) ΔNB. Task 1 No	(2) ΔNB. Task 1 Yes	(3) ΔNB. Task 2 No	(4) ΔNB. Task 2 Yes
Baseline = AI treatment dummy					
Expert treatment dummy		-0.266	0.833**	0.029	1.203***
		(0.253)	(0.348)	(0.148)	(0.291)
Student treatment dummy		0.010	1.641***	-0.133	1.021***
		(0.203)	(0.368)	(0.140)	(0.354)
Male dummy		-0.004	-0.393	0.096	-0.244
		(0.204)	(0.337)	(0.120)	(0.290)
Undergraduate student dummy		-0.032	0.394	-0.111	0.139
		(0.186)	(0.308)	(0.113)	(0.302)
Science major dummy		-0.158	-0.310	-0.125	-0.013
		(0.232)	(0.389)	(0.139)	(0.316)
Stock trading experience dummy		0.059	-0.130	0.059	0.308
		(0.298)	(0.490)	(0.178)	(0.419)
Financial literacy score		0.021	-0.015	-0.030	-0.002
		(0.042)	(0.070)	(0.031)	(0.063)
Constant		0.447	-2.546***	0.454	-3.128***
		(0.475)	(0.646)	(0.315)	(0.618)
Observations		1,851	1,229	1,832	1,248
Clusters		296	252	302	259
R-squared		0.002	0.029	0.002	0.020
Treatments		Predicted ΔNB. Task 1	Predicted ΔNB. Task 2	Predicted ΔNB. Task 1	Predicted ΔNB. Task 2
AI		0.489	-2.926	0.123	-3.178
		(0.161)	(0.227)	(0.113)	(0.217)
Experts		0.223	-2.094	0.152	-1.975
		(0.186)	(0.263)	(0.093)	(0.190)
Students		0.498	-1.285	-0.010	-2.157
		(0.118)	(0.277)	(0.086)	(0.279)
		Prob > F	Prob > F	Prob > F	Prob > F
AI = Experts = Students		0.421	0.000	0.383	0.000
AI = Expert #			0.052		0.000
AI = Student #			0.000		0.013
Expert = Student #			0.0997		1.000

Notes: (a) We regressed the change in net benefit (Part 2 - Part 1) (ΔNB.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table B7: Change in net benefit (points) when comparing with and without receiving advice conditional on personal characteristics

VARIABLES	(1) ΔNB. Task 1 Algorithm	(2) ΔNB. Task 1 Expert	(3) ΔNB. Task 1 Student	(4) ΔNB. Task 2 Algorithm	(5) ΔNB. Task 2 Expert	(6) ΔNB. Task 2 Student
Advice received dummy	-3.495***	-2.335***	-1.741***	-3.343***	-2.111***	-2.129***
	(0.281)	(0.311)	(0.281)	(0.252)	(0.190)	(0.292)
Male dummy	-0.067	-0.124	-0.334	0.104	-0.096	-0.074
	(0.345)	(0.328)	(0.256)	(0.252)	(0.239)	(0.236)
Undergraduate student dummy	0.494*	0.305	-0.585**	-0.322	0.168	0.085
	(0.280)	(0.310)	(0.238)	(0.228)	(0.229)	(0.291)
Science major dummy	-0.392	-0.534	0.384	-0.458	0.027	-0.151
	(0.397)	(0.424)	(0.272)	(0.281)	(0.269)	(0.251)
Stock trading experience dummy	0.341	-0.073	-0.650	-0.338	0.833**	0.052
	(0.384)	(0.515)	(0.521)	(0.305)	(0.392)	(0.350)
Financial literacy score	-0.042	0.030	0.016	-0.048	-0.028	0.021
	(0.074)	(0.076)	(0.050)	(0.052)	(0.064)	(0.059)
Constant	0.848	0.280	0.761	1.044**	0.222	-0.107
	(0.787)	(0.765)	(0.497)	(0.505)	(0.604)	(0.636)
Observations	1,060	1,030	990	1,060	1,030	990
Clusters	106	103	99	106	103	99
R-squared	0.141	0.080	0.060	0.181	0.118	0.105

Notes: (a) We regressed the change in net benefit (Part 2 - Part 1) (ΔNB.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecast with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses.

Table B8: Changes in WTP from Task 1 to Task 2 conditional on personal characteristics

VARIABLES	(1) Algorithm △WTP	(2) Expert △WTP	(3) Student △WTP	(4) Algorithm △WTP	(5) Expert △WTP	(6) Student △WTP
△Mean performance in Task 1	0.331***	0.213***	-0.031			
	(0.101)	(0.072)	(0.106)			
Average cost per advice paid in Task 1	-0.841***	-1.155***	-0.591***			
	(0.150)	(0.232)	(0.212)			
△Net benefit in Task 1				0.449***	0.455***	0.181*
				(0.094)	(0.116)	(0.108)
Male dummy	-0.174	-0.030	-0.032	-0.181	-0.268	0.260
	(0.336)	(0.355)	(0.439)	(0.335)	(0.366)	(0.470)
Undergraduate student dummy	-0.763**	-0.022	-0.790*	-0.851**	0.012	-0.544
	(0.328)	(0.446)	(0.464)	(0.356)	(0.579)	(0.466)
Science major dummy	-0.541	-0.378	-0.457	-0.500	0.008	-0.555
	(0.406)	(0.496)	(0.442)	(0.419)	(0.582)	(0.493)
Stock trading experience dummy	-0.136	0.054	-1.285	-0.240	0.094	-1.680
	(0.481)	(0.442)	(0.966)	(0.479)	(0.412)	(1.134)
Financial literacy score	-0.049	-0.086	-0.101	-0.016	-0.049	-0.137
	(0.082)	(0.103)	(0.079)	(0.083)	(0.131)	(0.088)
Constant	3.587***	3.673***	2.824**	2.260**	1.059	1.782*
	(0.974)	(1.186)	(1.160)	(0.936)	(1.202)	(0.998)
Obs.	97	84	71	97	84	71
R^2	0.378	0.455	0.323	0.303	0.257	0.189

Notes: a) In models 1-3, we regressed △WTP on △Mean performance from graphs with advice using an OLS regression model, controlling for the average cost per advice paid in Task 1 and personal characteristics, with robust standard errors. △WTP is calculated as WTP in Task 2 minus WTP in Task 1. △Mean performance is calculated as the mean performance from graphs with advice in Part 2 of Task 1 minus the mean performance from graphs with advice in Part 1 of Task 1. Performance ranges from 0 to 20, based on the points obtained in each question. In models 4-6, we regressed △WTP on △Net benefit from graphs with advice using an OLS regression model controlling personal characteristics, with robust standard errors. △Net benefit is calculated as the mean net benefit from graphs with advice in Part 2 of Task 1 minus the mean net benefit from graphs with advice in Part 1 of Task 1. The net benefit in each graph in Part 1 equals the rewards in each graph plus 10 points. The net benefit in each graph in Part 2 equals the rewards in each graph plus 10 points minus the price of advice. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.
b) The unit of observation is the participant who received at least one piece of advice in Task 1. The total number of observations is the number of participants who received at least one piece of advice in Task 1 in each treatment.
c) Numbers in parentheses represent standard errors.

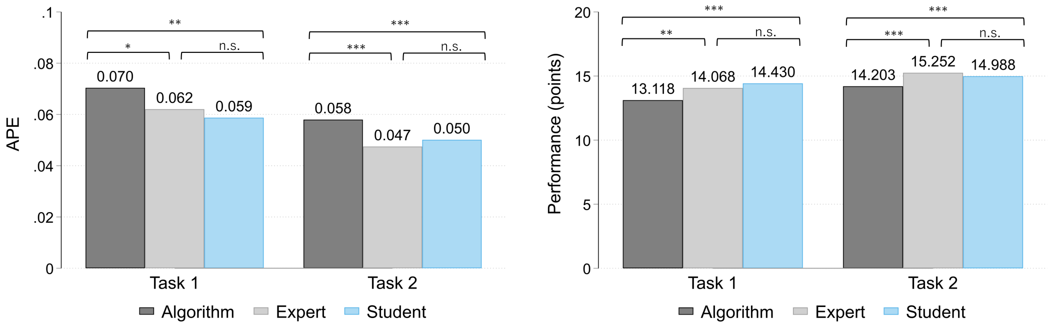
Table B9: Price of Advice and Shift Rate conditional on personal characteristics

VARIABLES	(1) Algorithm SHIFT	(2) Expert SHIFT	(3) Student SHIFT	(4) Algorithm SHIFT	(5) Expert SHIFT	(6) Student SHIFT
Price	0.000	0.004	0.003	-0.003	0.005	-0.005
	(0.005)	(0.006)	(0.009)	(0.004)	(0.005)	(0.010)
Task 2 dummy	0.006	-0.050*	-0.080**	0.006	-0.050*	-0.078**
	(0.021)	(0.026)	(0.034)	(0.020)	(0.026)	(0.033)
WTP				0.006	-0.002	0.013
				(0.008)	(0.012)	(0.016)
Male dummy	-0.045	-0.051	0.018	-0.043	-0.050	0.017
	(0.043)	(0.044)	(0.050)	(0.043)	(0.043)	(0.049)
Undergraduate student dummy	0.028	0.065	-0.020	0.028	0.064	-0.018
	(0.037)	(0.046)	(0.048)	(0.037)	(0.045)	(0.047)
Science major dummy	0.031	0.021	0.035	0.029	0.019	0.043
	(0.045)	(0.052)	(0.044)	(0.045)	(0.053)	(0.045)
Stock trading experience dummy	0.082*	0.058	0.007	0.079*	0.056	0.011
	(0.046)	(0.051)	(0.072)	(0.045)	(0.052)	(0.070)
Financial literacy score	-0.006	-0.002	-0.004	-0.006	-0.003	-0.005
	(0.010)	(0.011)	(0.012)	(0.010)	(0.011)	(0.011)
Constant	0.840***	0.822***	0.577***	0.813***	0.836***	0.539***
	(0.094)	(0.101)	(0.104)	(0.101)	(0.126)	(0.121)
Obs.	937	783	506	937	783	506
Clusters	101	93	78	101	93	78
R^2	0.021	0.028	0.021	0.022	0.028	0.025

Notes: a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between 0 and 1. We regressed shift rate on price of advice in each graph, using an OLS regression model, controlling for the Task 2 dummy , WTP and personal characteristics, with robust standard errors clustered at participant level. Price indicates the random price of each graphs in each task. Task 2 dummy equals to 1 for Task 2 and 0 for Task 1. WTP indicates the submitted WTP in each task. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.
b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in Task 1 and Task 2 in each treatment.
c) Numbers in parentheses represent standard errors.

C Other results

C.1 Performance of advice received by participants



(a) Absolute Percentage Error (APE)

(b) Performance (points)

Figure C1: Performance of advice received by participants

Notes: We calculated the average performance of the advice received by participants in each task across the three treatments, with one observation in Task 1 and one observation in Task 2 for each participant. We regressed (a) the APE of received advice and (b) the performance of received advice in points on three treatment dummy variable using an OLS regression model with robust standard errors. Pairwise comparisons of the estimated dummy coefficients were performed using an F-test with Bonferroni-adjusted p-values. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, and n.s. denotes that the difference is not statistically significant at the 0.1 level.

Figure C1 illustrates the average performance of the advice received by participants across the three treatments. In both tasks, the average performance of the received advice, in terms of APE and performance points, is worse in the algorithm treatment compared to the expert treatment and the student treatment, and is similar between the expert treatment and the student treatment.

C.2 Change of APE

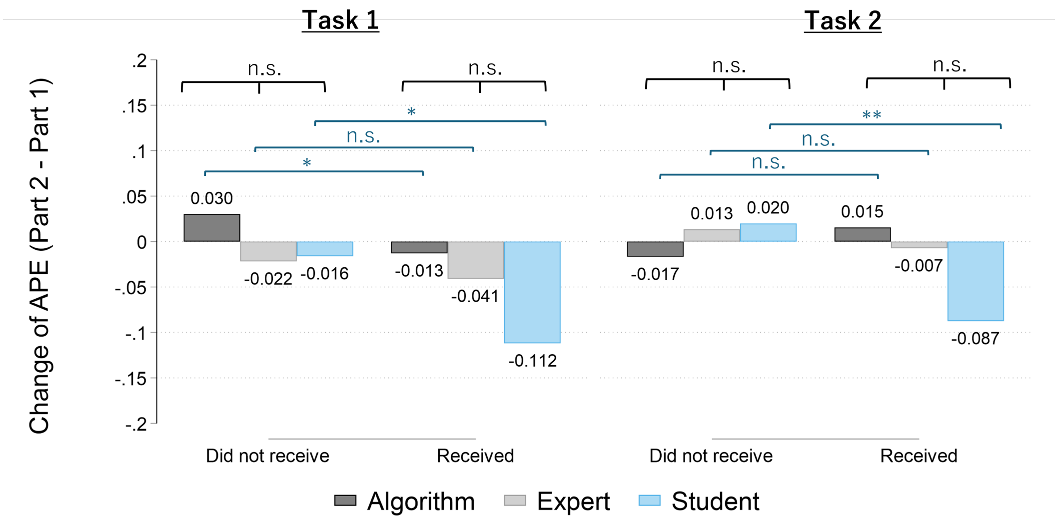


Figure C2: Change of APE

Notes: Outliers (change of APE ≤ -20 or ≥ 20) are removed in this analysis, resulting in the removal of three observations. We regressed the change in APE (Part 2 - Part 1) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. Pairwise comparisons of the estimated dummy coefficients were performed using an F-test with Bonferroni-adjusted p-values. We also regressed the change in performance points (Part 2 - Part 1) on an advice-receiving dummy in each treatment using an OLS regression model with robust standard errors clustered at the participant level. The symbols *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, while n.s. denotes that the difference is not statistically significant at the 0.1 level.

Figure C2 illustrates the average change in absolute percentage error (APE) from Part 1 to Part 2 in Task 1 (left) and Task 2 (right) for each of the three treatments. In both tasks, the performance improvement from Part 1 to Part 2 among participants who received advice is better than that of participants who did not receive advice (except for those in the algorithm treatment in Task 2). These results are significant for advice from algorithms and students in Task 1, and for advice from students in Task 2.

C.3 WTP vs. Net benefit in Part 2

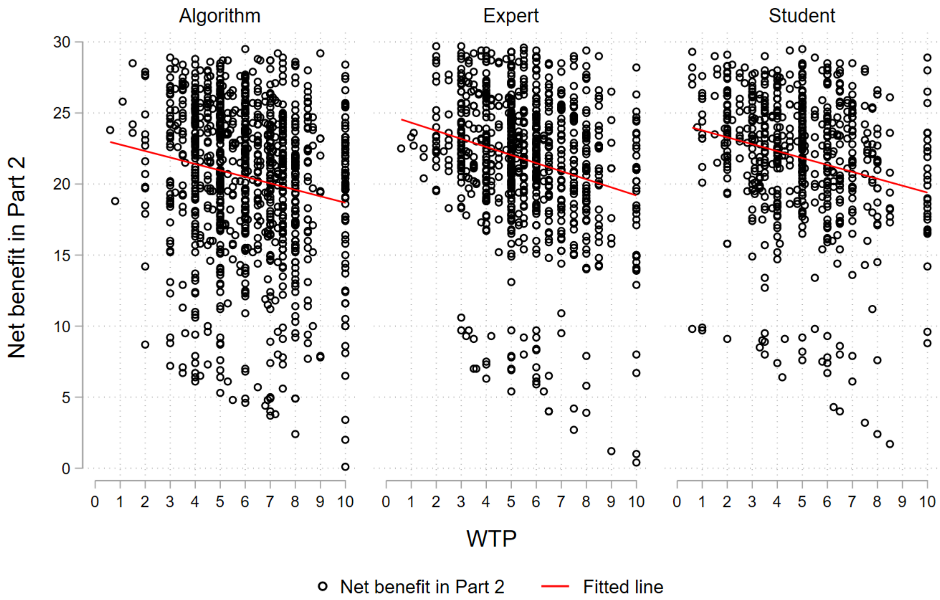


Figure C3: WTP vs. Change of performance (points)

Figure C3 presents a scatter plot of WTP values and their corresponding net benefits in Part 2 for both tasks, measured in points (Part 2 - Part 1), across treatments, incorporating data from both Task 1 and Task 2. The red line represents the linear fit of net benefits in Part 2 (points) on WTP. Across all treatments, net benefits in Part 2 decrease as WTP increases, after paying and receiving for the advice..

C.4 WTP vs. Change of performance (points)

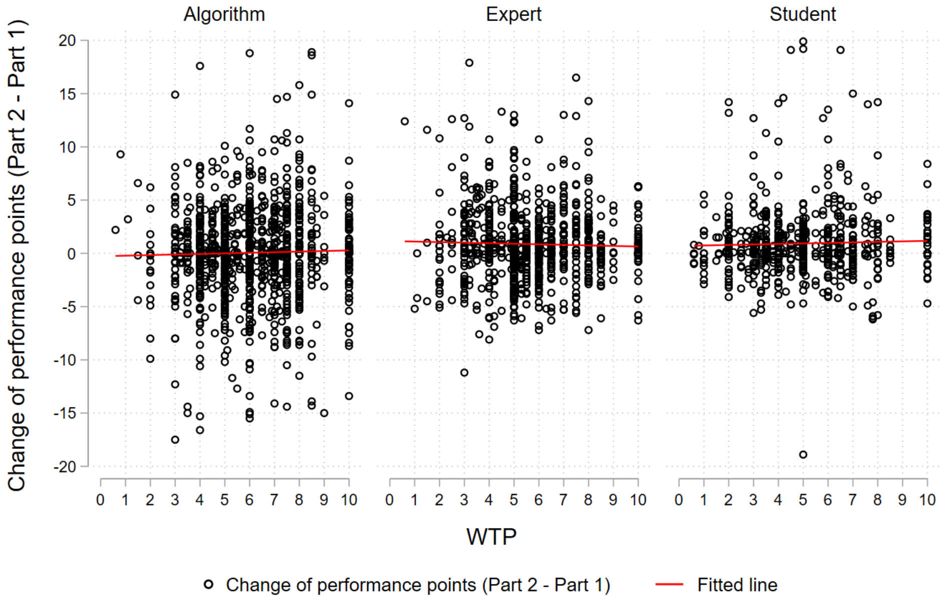


Figure C4: WTP vs. Change of performance (points)

Figure C4 presents a scatter plot of WTP values and their corresponding changes in performance, measured in points (Part 2 - Part 1), across treatments, incorporating data from both Task 1 and Task 2. The red line represents the linear fit of the change in performance (points) on WTP. To estimate this relationship, we regressed the change in performance (points) on WTP using an OLS regression model with robust standard errors clustered at the participant level. Based on this model, we identified the WTP value at which the change in performance (points) equals zero.

The minimum WTP required to achieve an increase in performance (points) was 4.958 for the algorithm. For both the expert and student treatments, all observed WTP values resulted in an increase in performance (points).

D Experimental Instruction (English Translation)

Experimental Instruction

1

Welcome to the experiment

- We will now begin the experiment. In addition to the participation fee of 500 yen for taking part in this experiment, you will also receive reward based on the choices you make during the experiment.
- You are not allowed to communicate with other participants during the experiment. Additionally, please turn off all electronic devices such as smartphones and iPads during the experiment.
- If you have any questions during the experiment, please raise your hand. We will answer them individually.

2

Experiment details

The experiment consists of three parts:

Part 1: Questionnaire
Part 2: Experiment Task 1
Part 3: Experiment Task 2

The total duration of the experiment is approximately one hour.

3

Experiment details: Questionnaire

- In the questionnaire, you will be asked mainly about your experience with stock investments and your knowledge of financial literacy to assess your knowledge and experience related to financial investment.
- Your responses will not affect your reward, but please answer honestly according to your personal situation.
- The estimated time for completing the questionnaire is 10 minutes.
- If the time limit is exceeded, you will be automatically moved to the next step.
- Once you have completed the questionnaire, click the 'Next' button to begin Task 1.

4

Experiment details: Tasks 1 & 2

- In Tasks 1 and 2, you will be asked to **predict future stock prices based on past stock price information.** Your reward will depend on how accurate your predictions are, as will be explained later.
- In each question, you will be shown a graph displaying the closing prices of a particular stock over a 12-month period. You will then be asked to predict the stock price 30 days after the last day shown on the graph.
- The period for each graph is randomly selected from the time between January 1, 2008, and June 30, 2018, and will cover a 12-month period starting from that randomly selected date.
- The stock will be randomly chosen from the S&P 500.
- The stock name and the exact period will not be displayed.

5

Experiment details: Tasks 1 & 2

- Tasks 1 and 2 will follow the same procedure.
- Each task is divided into two parts: the first half and the second half.

Task 1	Task 2	Steps
First half	First half	After reading the graph of a certain stock, make a prediction.
		Submit your willingness to pay for additional information.
Second half	Second half	After reading the graph and additional information, make a prediction.

Same 10 graphs   Same 10 graphs   Total 20 graphs

- More details will be provided later.

6

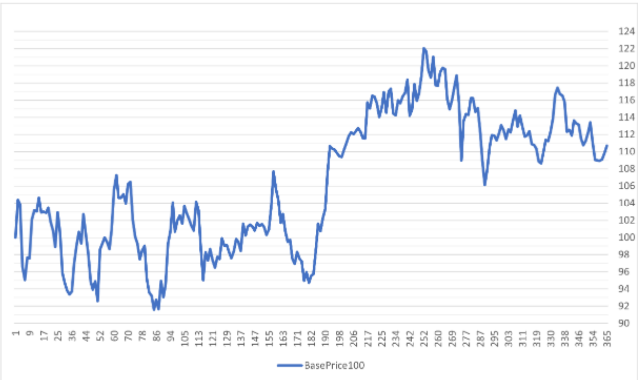
Experiment details：The first half of Tasks 1 & 2

- In the first half, as mentioned earlier, you will read a graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph.
- All graphs are normalized so that the stock price on the first day of the graph is set to 100.
- Please submit your predicted value with up to three decimal places.
- There are 10 questions in total.
- **<u>The time limit for each question is 40 seconds.</u>**
- In the next slide, a sample question will be displayed.

7

Sample Question

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 110.693.



Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

8

Experiment details：The second half of Tasks 1 & 2

- In the second half, the same 10 graphs will be displayed in the same order, and you will make predictions again. The time limit for each question is 40 seconds.
- However, in the second half, you may have the opportunity to obtain additional information to assist with your predictions.
- Whether you can obtain the additional information depends on the amount you submit between the first and second halves as **your willingness to pay** (WTP, the maximum amount you are willing to pay to purchase the additional information) and the **randomly determined price** (p) of the additional information.
- This will be explained in more detail later.

9

Experiment details：The second half of Tasks 1 & 2

- **Participants who obtained additional information**: they can refer to both the graph and the additional information to predict the stock price again.
- **Participants who did not obtain additional information**: they will predict the stock price again using only the graph, just as in the first half.
- As in the first half, please submit your predicted value with up to three decimal places.
- There are 10 graphs in total.

10

Obtaining additional information

- Before the second half begins, an explanation will be provided on how the availability of additional information is determined, based on the submitted willingness to pay (WTP) and the price (p) of the additional information, which is randomly set for each question.
- In this experiment, the WTP that can be submitted is limited to numbers **from 0.0 to 10.0, with one decimal place**.
- The price of the additional information (p) ranges **from 0.1 to 10.0 and is randomly determined** for each of the 10 questions in the second half. All prices are chosen with the same probability.

11

Purchase of additional information

- In a given graph, if the WTP is greater than or equal to the price of the additional information (p) (WTP ≥ p):
  - You will pay p and obtain the additional information.
- Conversely, if the WTP is less than p (WTP < p):
  - You will not obtain the additional information, and no payment will occur.
- The WTP in all 10 questions is the same and is submitted only once before the second half begins.
- The price of the additional information (p) is determined randomly for each of the 10 questions in the second half.

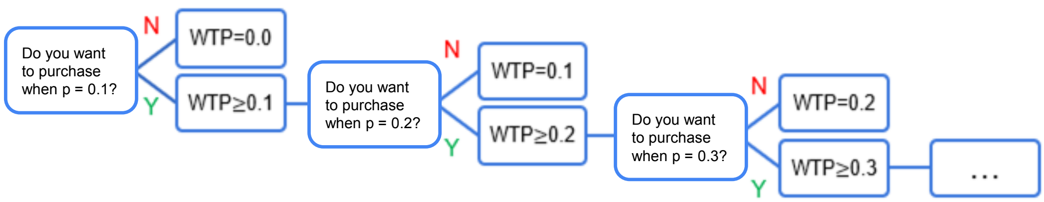
12

Purchase of additional information

- Example 1
  - For instance, suppose you submit a WTP of 10.0.
  - In this case, since the price of the additional information determined randomly for each graph is at most 10.0, you will obtain additional information for all graphs and pay the price of the additional information for each graph.
- Example 2
  - For instance, suppose you submit a WTP of 0.0.
  - In this case, since the minimum price of the additional information determined randomly for each graph is 0.1, you will not obtain additional information for any graph, and no payment for the additional information will occur either.

How to determine the willingness to pay (WTP)

When determining the WTP, first consider the following questions:



Continue this until your answer changes from 'yes' to 'no'.
The price at which your answer changes is your WTP.

Endowment for purchasing additional information

- For Task 1 and Task 2, you are provided with 100 points endowment.
- Out of these 100 points, any points not used for purchasing additional information will be added to your rewards at a rate of 1 point = 6 yen.
- Example 1: If you submit a WTP of 0.0 and do not purchase any additional information, these 100 points will be added to your rewards.
- Example 2: If you submit a WTP of 10.0 and obtain additional information for each of the 10 problems, spending a total of 90 points, the remaining 10 points will be added to your rewards.

15

Example of Purchasing Additional Information and Point Calculation (WTP = 5.0)

Question	Price of additional information(p)	Can the additional information be purchased?	Remaining points
Q1	3.5	Yes	100-3.5=96.5
Q2	6.2	No	96.5
Q3	0.3	Yes	96.5-0.3=96.2
Q4	7.3	No	96.2
Q5	8.2	No	96.2
Q6	5.0	Yes	96.2-5.0=91.2
Q7	3.8	Yes	91.2-3.8=87.4
Q8	5.1	No	87.4
Q9	5.0	Yes	87.4-5.0=82.4
Q10	9.2	No	82.4

16

Calculation of rewards

- In addition to a participation fee of JPY 500, you can earn additional rewards based on your performance in Tasks 1 and 2.
- One of the four parts—Task 1 first half, Task 1 second half, Task 2 first half, and Task 2 second half—will be randomly selected, and additional rewards will be determined based on the accuracy of the 10 predictions you submitted.
- Please refer to the table below for the calculation method of the additional rewards for each prediction:

Degree of deviation from the true value	Points
±0%	20
±0% ~±20%	20-|Degree of deviation×100|
±20%~	0

- The total points gained from the 10 questions is used to calculate the additional reward.
- The points earned are converted to Japanese yen at a rate of 1 point = 6 yen.

17

Example of point calculation

For example, suppose the true value for a particular question 118.890.

- If your prediction is 130.123：

The degree of deviation from the true value is

$\left|\frac{130.123-118.890}{118.890}\right| \approx 0.09448\ (9.448\%)$

You get $20-|0.09448 \times 100| \approx 10.552$ points.

- If your prediction is 90.002：

The degree of deviation from the true value is

$\left|\frac{90.002-118.890}{118.890}\right| \approx 0.243\ (24.3\%)$

You get no points.

18

Summary regarding rewards

In today's experiment, in addition to a participation fee of 500 yen, the total rewards paid to you will include:

- The amount based on the points not used for purchasing additional information, from the 100 points provided for purchasing additional information in Task 1 and Task 2.
- The amount based on the points earned according to the accuracy of the 10 predictions in one of the four parts—Task 1 first half, Task 1 second half, Task 2 first half, Task 2 second half—randomly selected for rewards.

This total amount will be paid to you as total rewards.

19

This concludes the explanation of the experiment. Please answer the quiz to confirm whether you have correctly understood the content of the explanation.

20

E Quiz (English Translation)

1. What can you do in this experiment?

   1 View the stock price graph of a company you specify

   2 Submit your predictions

   3 Decide the part used for calculating additional rewards

   4 Choose the source of additional information you want to obtain

   [Correct answer: 2]

2. In which of the following cases do you obtain additional information?

Case	Willingness to pay (WTP)	Price of additional information (p)
1	10	5
2	4	6
3	8	8
4	3	9

   1 Case 1 and Case 3

   2 Case 2 and Case 4

   3 Case 1 and Case 4

   4 Case 2 and Case 3

   [Correct answer: 1]

3. Which of the following is true about the willingness to pay (WTP)?

   1 It is possible to submit the WTP 10 times per task.

   2 It is possible to submit 12.

   3 It is not possible to submit 5.5.

   4 It is the maximum number of points you are willing to pay to acquire additional information.

   [Correct answer: 4]

4. Which of the following is true about the price of additional information (p)?

   1 p = 12 is possible.

   2 There is only one per task.

   3 It is determined randomly for each question in the second half of the task.

   4 You can set it yourself.

[Correct answer: 3]

5. Please select the correct option regarding the purchase of additional information.

   1 If WTP $\geq$ p, you decide whether to buy it yourself.

   2 If WTP $<$ p, you cannot purchase, but you will still pay the points.

   3 If WTP $=$ p, you cannot purchase.

   4 If WTP $\geq$ p, you automatically pay p points.

   [Correct answer: 4]

F Questionnaire (English Translation)

F.1 Survey on demographic characteristics

1. Gender:

   ☐ Female

   ☐ Male

   ☐ Other

   ☐ Do not want to answer

2. Grade level:

   ☐ Undergraduate Year 1

   ☐ Undergraduate Year 2

   ☐ Undergraduate Year 3

   ☐ Undergraduate Year 4

   ☐ Master Year 1

   ☐ Master Year 2

   ☐ Other

3. Major:

   ☐ Humanities

   ☐ Sciences

F.2 Survey on stock trading experience

1. Do you have experience in stock investment?

   ☐ Yes (Go to Q2 - Q6)

   ☐ No

2. How much experience did you have with stock investments during your student days?

   ☐ Less than six months

   ☐ Six months to one year

   ☐ One to three years

   ☐ More than three years

3. What is the main type of stock you target in your investments?

   ☐ Large-cap stocks

   ☐ Small-cap stocks

   ☐ Technology stocks

   ☐ Consumer goods stocks

   ☐ Financial stocks

   ☐ Other

4. What is the performance of your investments?

   ☐ Stable profits

   ☐ Stable losses

   ☐ Balanced between profits and losses

   ☐ Uncertain with no clear trend

5. What are the main sources of information you use when investing in stocks?

   ☐ Financial news and media coverage

   ☐ Corporate financial reports and fundamental analysis

   ☐ Technical analysis and chart patterns

   ☐ Social media and forum discussions

   ☐ Professional research reports

   ☐ Other

6. What is your perception of the importance of risk management in stock investing?

☐ Extremely important, and I actively employ risk management strategies.

☐ Important, but there may be room for improvement.

☐ Not very emphasized, focusing more on profits.

☐ I do not understand the concept of risk management at all.

F.3 Financial literacy test

The financial literacy scores were measured by following Fernandes et al. (2014). The 12 questions were as follows.

1. Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy:

   1 More than today with the money in this account.

   2 Exactly the same as today with the money in this account.

   3 Less than today with the money in this account.

   4 Don't know.

   5 Refuse to answer.

   [Correct answer: 3]

2. Do you think that the following statement is true or false? "Bonds are normally riskier than stocks."

   1 True.

   2 False.

   3 Don't know.

   4 Refuse to answer.

   [Correct answer: 2]

3. Considering a long time period (for example 10 or 20 years), which asset described below normally gives the highest return?

   1 Savings accounts.

   2 Stocks.

   3 Bonds.

   4 Don't know.

   5 Refuse to answer.

   [Correct answer:2]

4. Normally, which asset described below displays the highest fluctuations over time?

   1 Savings accounts.

   2 Stocks.

   3 Bonds.

   4 Don't know.

   5 Refuse to answer.

   [Correct answer:2]

5. When an investor spreads his money among different assets, does the risk of losing a lot of money:

   1 Increase.

   2 Decrease.

   3 Stay the same.

   4 Don't know.

   5 Refuse to answer.

   [Correct answer:2]

6. Do you think that the following statement is true or false? "If you were to invest $1000 in a stock mutual fund, it would be possible to have less than $1000 when you withdraw your money?"

   1 Increase.

   2 Decrease.

   3 Stay the same.

   4 Don't know.

   5 Refuse to answer.

   [Correct answer: 1]

7. Do you think that the following statement is true or false? "A stock mutual fund combines the money of many investors to buy a variety of stocks."

   1 True.

   2 False.

   3 Don't know.

   4 Refuse to answer.

[Correct answer: 1]

8. Do you think that the following statement is true or false? “A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less.”

   1 True.

   2 False.

   3 Don’t know.

   4 Refuse to answer.

   [Correct answer: 1]

9. Suppose you had \$100 in a savings account and the interest rate is 20% per year and you never withdraw money or interest payments. After 5 years, how much would you have on this account in total?

   1 More than \$200.

   2 Exactly \$200.

   3 Less than \$200.

   4 Don’t know.

   5 Refuse to answer.

   [Correct answer: 1]

10. Which of the following statements is correct?

    1 Once one invests in a mutual fund, one cannot withdraw the money in the first year.

    2 Mutual funds can invest in several assets, for example invest in both stocks and bonds.

    3 Mutual funds pay a guaranteed rate of return which depends on their past performance.

    4 None of the above

    5 Don’t know.

    6 Refuse to answer.

    [Correct answer: 2]

11. Which of the following statements is correct? If somebody buys a bond of firm B:

    1 He owns a part of firm B.

2 He has lent money to firm B.

3 He is liable for firm B's debts.

4 None of the above

5 Don't know.

6 Refuse to answer.

[Correct answer: 2]

12. Suppose you owe \$3,000 on your credit card. You pay a minimum payment of \$30 each month. At an Annual Percentage Rate of 12% (or 1% per month), how many years would it take to eliminate your credit card debt if you made no additional new charges?

    1 Less than 5 years.

    2 Between 5 and 10 years.

    3 Between 10 and 15 years.

    4 None of the above

    5 Don't know.

    6 Refuse to answer.

    [Correct answer: 4]

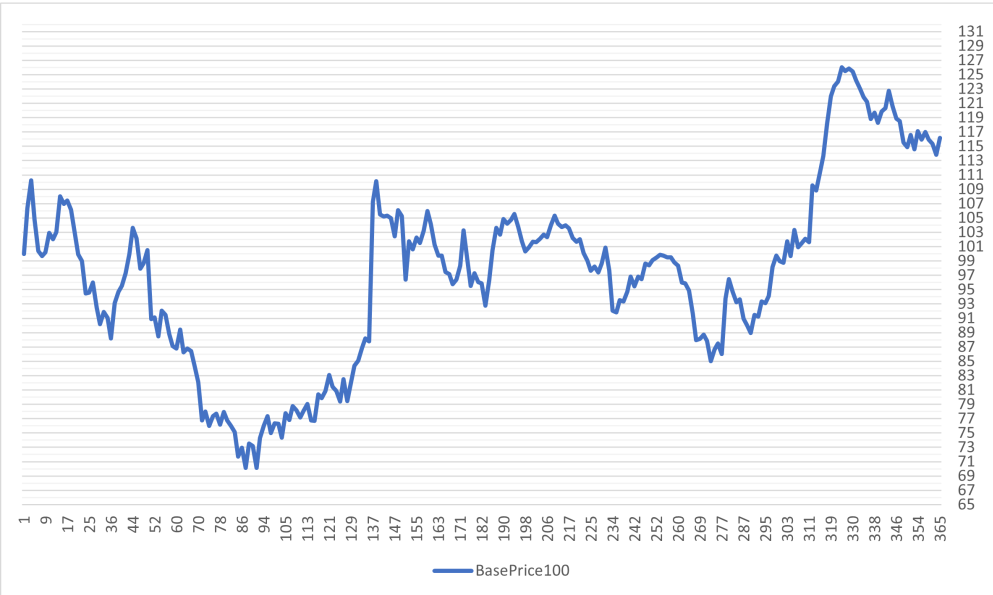
G Screenshots of Tasks 1 and 2 (English Translation)

[Screen 1] There are 10 questions in Part 1 of both Task 1 and Task 2. Below is an example of Question 1 from Part 1 of Task 1.

Question 1

The remaining time: 0:40

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 116.183.



Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

[Screen 2]

Question 1

The remaining time: 0:10

Your prediction has been successfully sent.
Your prediction for Question 1 is 118.3.

Next

[Screen 3] (shown in algorithm treatment)

Information source and Willingness to pay (WTP)

The remaining time: 0:60

The additional information provided here is a stock price forecast submitted based on the same stock price chart used in this experiment, by an algorithm crafted to predict future stock prices.

This algorithm makes the future stock price forecast by learning the historical stock price information, from January 1st, 2000 or from the Initial Public Offering (IPO) day to January 1st, 2020, of 83 target companies rank top in their capital market sectors (i.e. Basic Materials, Consumer Goods, Healthcare, Services, Utilities, Conglomerates, Financial, Industrial Goods, Technology).

Now your endowment is 100 points.

How much are you willing to pay, up to a maximum of 10.0, for information from the algorithm?

Note: WTP submissions are once per task only.
Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 3] (shown in expert treatment)

Information source and Willingness to pay (WTP)

The remaining time: 0:60

The additional information provided here is the average of stock price predictions submitted by 198 experts (CMAs) who participated in a similar experiment in the past, based on the stock price chart used in this experiment, for the stock prices 30 days ahead.

CMA stands for Certified Member Analyst of the Securities Analysts Association of Japan. The CMA is a qualification granted to those who have taken the prescribed training courses and passed the examinations based on these courses and fulfilled certain requirements, and is a sign of a expert in the fields of finance and investment. The CMA's investment valuation is based on the calculation of the corporate value of an investment and the forecasting of its future value. Therefore, the particularly crucial knowledge can be broadly categorized into three main areas: evaluation of investment data, decision-making on investment policies, and construction and management of portfolios.

Now your endowment is 100 points.

How much are you willing to pay, up to a maximum of 10.0, for information from the algorithm?

Note: WTP submissions are once per task only.
Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 3] (shown in student treatment)

Information source and Willingness to pay (WTP)

The remaining time: 0:60

The additional information provided here is the average of stock price predictions submitted by 233 Osaka University students who participated in a similar experiment conducted in the past, based on the stock price chart used in this experiment, for the stock prices 30 days ahead.

Now your endowment is 100 points.

How much are you willing to pay, up to a maximum of 10.0, for information from the algorithm?

Note: WTP submissions are once per task only.
Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 4] (shown in the case WTP $\geq$ p)

Notice

The remaining time: 0:15

The Price of additional information is 1.1.
Your willingness to pay is 4.0.

Because WTP $\geq$ p, you can obtain additional information in the next question.

Your current points is 98.9 points.

Next

[Screen 4] (shown in the case WTP $<$ p)

Notice

The remaining time: 0:15

The Price of additional information is 8.0.
Your willingness to pay is 4.0.

Because WTP < p, you cannot obtain additional information in the next question.

Your current points is 100.0 points.

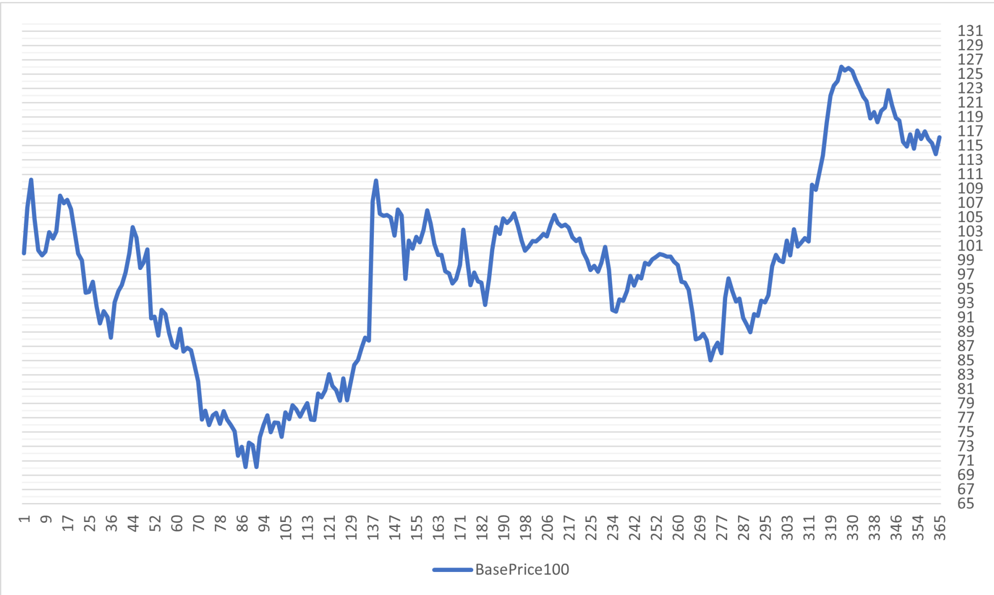
Next

[Screen 5] There are 10 questions in Part 2 of both Task 1 and Task 2. The order of questions in Part 2 is the same as in Part 1 in each task. Below is an example of Question 11 from Part 2 of Task 1.

Question 11

The remaining time: 0:40

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 116.183.



Additional information: The prediction value from the algorithm is 124.502. (shown in algorithm treatment and in the case WTP $\geq$ p)
Additional information: The prediction value from the experts is 115.984. (shown in

expert treatment and in the case WTP $\geq$ p)

Additional information: The prediction value from the Osaka University students is 113.969. (shown in student treatment and in the case WTP $\geq$ p)

Your prediction in first half is 118.3.

Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

[Screen 6] The feedback is shown at the end of Task 1 and Task 2. The following is an example of the feedback given at the end of Task 1.

Completion of Task 1 and Notification of Results

Thank you for your participation in Task 1.

We will inform you of the results of Task 1.
Total points in Part 1: 132.1 + 100 = 232.1 points
Total points in Part 2: 137.8 + 89.1 = 226.9 points

Please refer to the table below for the details of the gained points.
Click the next button to start Task 2.

Part 1

Questions	Your forecast	Points
1	118.3	10.1
2	140	8.4
3	113	16.8
4	144	15.9
5	100	12
6	143	0
7	133	19.7
8	101	18.6
9	107	17.1
10	123	13.5

Part 2

Questions	WTP	p	Your forecast	Points
11	4.0	1.1	124	4.8
12	4.0	9	140	8.4
13	4.0	5.7	112	17.7
14	4.0	2.3	140	18.8
15	4.0	6.2	107	18.5
16	4.0	9.4	144	0
17	4.0	9.4	133	19.7
18	4.0	5.1	101	18.6
19	4.0	3.8	104	14.4
20	4.0	3.7	119	16.9

Next

[Screen 7]

The end of experiment and reward announcement

Thank you for participating in the experiment today.
Please wait on this screen until everyone has finished.

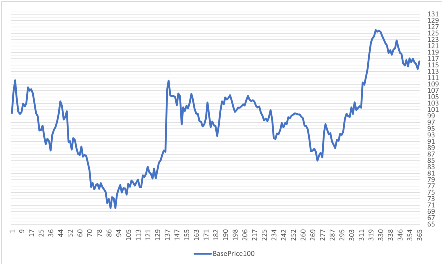
We will inform you about the rewards.

Task 1:
Points in the first half: 232.1 points
Points in the second half: 226.9 points

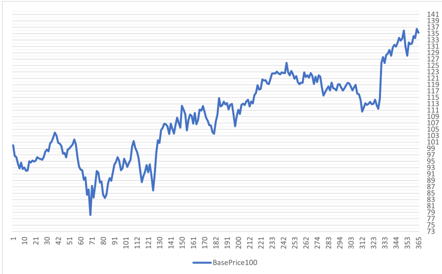
Task 2:
Points in the first half: 2252.6 points
Points in the second half: 235.5 points

The selected part for reward calculation is first half in Task 1.
The reward you will receive is calculated as 232.1 points $\times$ 6 JPY/point + JPY 500 = JPY 1892.6.

H Graphs in Task 1 and Task 2

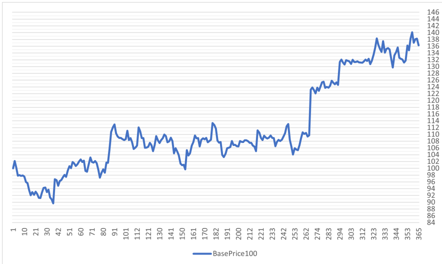


(a) Graph 1

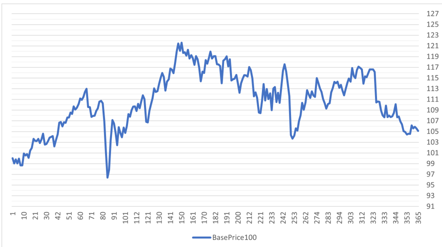


(b) Graph 2

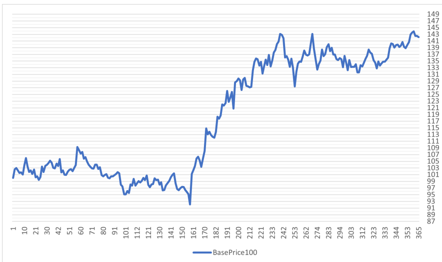
(c) Graph 3



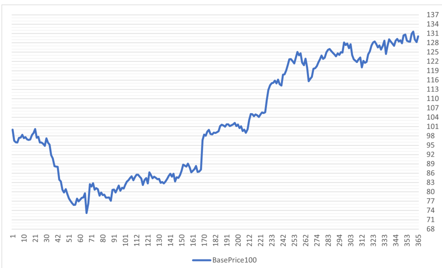
(d) Graph 4



(e) Graph 5

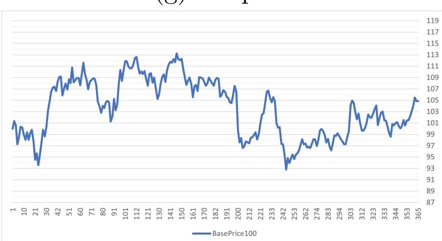


(f) Graph 6

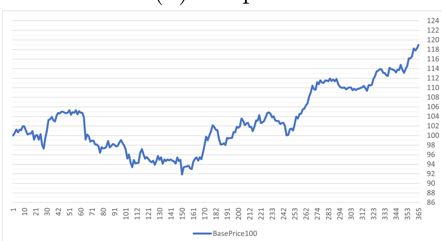


(g) Graph 7

(h) Graph 8

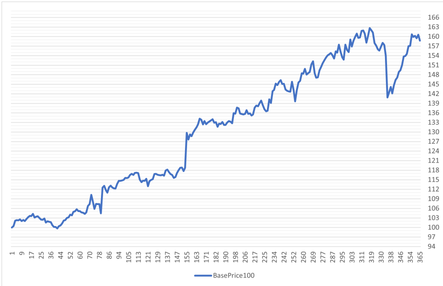


(i) Graph 9

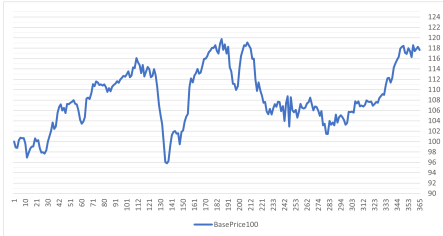


(j) Graph 10

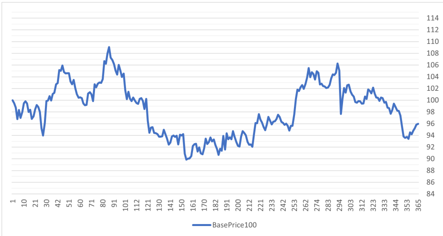
Figure H5: 10 graphs in Task 1



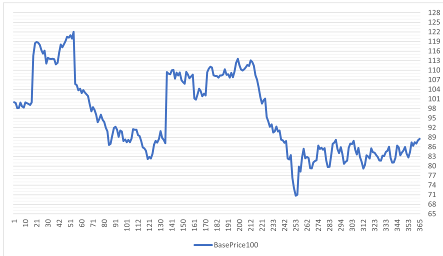
(a) Graph 11



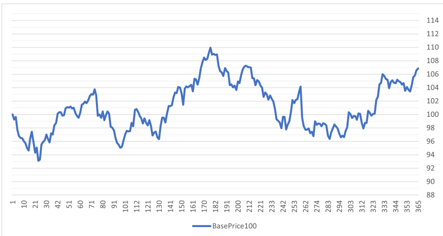
(b) Graph 12



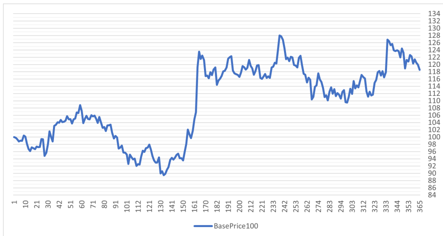
(c) Graph 13



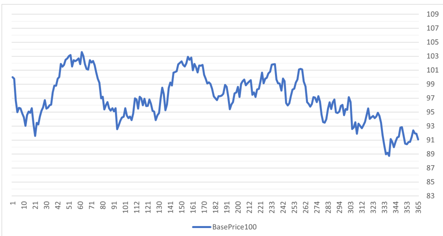
(d) Graph 14



(e) Graph 15

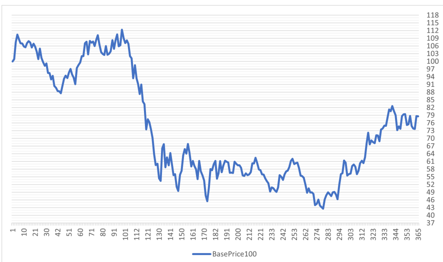


(f) Graph 16

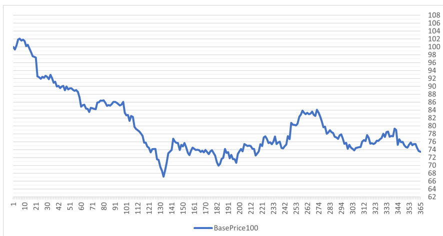


(g) Graph 17

(h) Graph 18



(i) Graph 19



(j) Graph 20

Figure H6: 10 graphs in Task 2

Table H1: Stock Data with algorithm, expert, and student advice

Graph	Stock Ticker	Period Date (Start from June)	Closing Price at Last Date (Normalized)	Closing Price One Month Later (Normalized)	AI Advice	Expert Advice	Student Advice
1	STX	2010-2011	116.183	107.676	124.502	115.984	113.969
2	LKQ	2011-2012	135.314	125.498	136.293	134.380	133.211
3	UPS	2009-2010	116.381	109.475	118.215	116.255	115.061
4	DLTR	2009-2010	136.241	138.273	139.922	137.484	135.960
5	SBUX	2015-2016	105.113	108.656	105.449	105.973	105.031
6	FOXA	2017-2018	142.146	183.223	143.164	141.515	139.942
7	TSN	2012-2013	130.098	132.576	130.824	130.918	130.610
8	COP	2009-2010	105.073	102.479	107.670	105.877	105.305
9	F	2013-2014	104.847	110.204	104.032	104.407	103.783
10	FISV	2016-2017	118.940	115.456	118.870	122.045	120.121
11	EL	2017-2018	158.742	151.572	162.390	157.382	157.686
12	ADM	2014-2015	117.637	108.027	119.225	113.552	114.655
13	NKE	2016-2017	95.962	96.288	97.026	96.434	96.796
14	TRIP	2015-2016	88.537	83.505	92.265	91.306	90.399
15	WM	2013-2014	106.869	106.940	109.501	107.334	110.741
16	MLM	2016-2017	118.546	117.107	122.650	116.721	118.174
17	HES	2013-2014	91.097	97.096	93.155	89.991	90.655
18	SIVB	2016-2017	130.597	136.070	134.396	130.488	129.888
19	SPGI	2008-2009	78.589	73.714	77.623	78.882	79.430
20	FMC	2014-2015	73.448	67.383	75.437	73.581	73.534