Discussion Paper No. 1156

ISSN (Print) 0473-453X
ISSN (Online) 2435-0982

FINANCIAL FORECASTING IN THE LAB AND THE FIELD: QUALIFIED PROFESSIONALS VS. SMART STUDENTS

Te Bao
Brice Corgnet
Nobuyuki Hanaki
Katsuhiko Okada
Yohanes E. Riyanto
Jiahua Zhu

Revised September 2024
January 2022

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

Financial Forecasting in the Lab and the Field: Qualified Professionals vs. Smart Students*

Te Bao†, Brice Corgnet‡, Nobuyuki Hanaki§, Katsuhiko Okada, Yohanes E. Riyanto††, Jiahua Zhu‡‡**

Abstract

We compare the performance of financial professionals (CFAs) with university students in four financial forecasting tasks ranging from simple lab prediction tasks to longitudinal field prediction tasks. Although students and professionals performed similarly in the most artificial forecasting tasks, CFAs outperformed students in the field predictions. Differences in forecasting performance between finance professionals and students were explained by financial literacy, not cognitive ability.

Keywords: Financial forecasting, financial professionals, financial literacy, cognitive skills.

JEL Codes: C91, C93, G17, D91, G41.

* The experiments reported in this paper have been approved by the IRB of the Institute of Social and Economic Research (ISER), Osaka University. We gratefully acknowledge financial support from grant-in-aid for scientific research (KAKENHI, grant number: 18K19954, 20H05631,23H00055) from the Japan Society for the Promotion of Science (JSPS) as well as support from the Joint Usage/Research Center at ISER, Osaka University. We thank the Securities Analysts Association of Japan for their support in recruiting participants for our experiment. We also thank Kazuhito Ogawa and staff members at the Research Institute of Socionetwork Strategies at Kansai University for their support in conducting our additional experiments. This research has been performed within the framework of LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) run by the French National Research Agency (ANR). Declaration of interests: None.

† Division of Economics and NTU-WeBank Joint Research Center, Nanyang Technological University. E-mail: baote@ntu.edu.sg
‡ Emlyon Business School. E-mail: corgnet@em-lyon.com
§ Institute of Social and Economic Research, Osaka University and University of Limassol. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp
** Institute of Business and Accounting, Kwansei Gakuin University, and Magne-Max Capital Management. E-mail: katsuokada@magne-max.com
†† Corresponding author. Division of Economics, Nanyang Technological University. E-mail: yeriyanto@ntu.edu.sg
‡‡ King's Business School, King's College London, United Kingdom. E-mail: jiahua.zhu@kcl.ac.uk

1. Introduction

Financial forecasting has occupied the mind of people for centuries, possibly starting with the use of Babylonian tablets to predict crop prices (Lo and Hasanhodzic, 2010). Prominent scientists in all fields have also tried their luck in forecasting stock prices, often with little success.[1] Financial forecasting likely appeals to the crowds not only because success can bring fortune, but also because any enduring recipe for success must remain secret. However, finance scholars have long argued that no winning recipe can exist because markets are informationally efficient, so all the relevant information for predicting stocks is already incorporated into prices (Samuelson, 1965; Fama, 1965, 1970, 1991; Batchelor, 1990; Hartzmark, 1991; Barber and Odean, 2000; Qu et al., 2019). They postulate that asset prices follow a random walk, and any successful prediction can only be due to luck (Malkiel, 1999; Tetlock, 2009b).

To study the extent of financial forecasting expertise, we conduct a series of forecasting experiments with financial professionals and compare their performance with that of students from highly selective (elite) and less selective (non-elite) universities. Our main research question is thus: *Will finance professionals outperform students in financial forecasting tasks?*

Our study contributes to the existing literature on financial forecasting, which relies on unincentivized surveys and does not provide direct comparisons between financial professionals and laypeople (see Cowles, 1933 for a historical example). Numerous studies use data from the Livingston Survey, a biannual longitudinal survey of experts' forecasts that started in 1946 and was commissioned by the Federal Reserve Bank of

[1] Isaac Newton is an often-cited example (see Malkiel, 1999, p. 45).

Philadelphia. A panel of experts consisting of US economists from the public and private sectors and academics is asked to provide stock market forecasts for six-month and twelve-month horizons. The last survey from December 2021 lists twenty-five experts. Livingston Survey studies have shown that the average expert does not perform better than the random walk model in predicting the US stock returns (Lakonishok, 1980; Pearce, 1984; De Bondt, 1991).

Furthermore, forecasts are often, though not always, biased (Brown and Maital, 1981; Dokko and Edelstein, 1989; Ackert et al., 1997, 2008a, 2008b) and miscalibrated (Deaves, et al., 2010; Ben-David et al., 2013; Boutros et al., 2020; Graham, 2022). Ackert et al. (1997) show that analysts' forecasts are not always consistent with their private information. Ben-David et al. (2013) find that, although financial officers made unbiased predictions for the return of the S&P 500 index on average, their confidence intervals were too narrow, thus exhibiting substantial miscalibration that persisted over the years of the survey (Boutros et al., 2020). Anecdotally, Kahneman (2011) also reports, using archival data, that the performance ranking of financial advisors is not correlated across years. All these findings echo the work of Tetlock (2009a), who showed the inability of experts to predict major political events consistently and likened forecasting to tarot card reading (Tetlock, 2009b).

Despite the early and numerous claims that there is no expertise in financial forecasting, a number of factors spark skepticism. At the theoretical level, the argument of finance scholars relies on the belief in the efficiency of markets, but this argument has been challenged by the advent of behavioral finance, which shows that psychological biases and heuristics often produce inefficiencies (De Bondt and Thaler, 1985, 1987; Lo and MacKinlay, 1988; Bernard and Thomas, 1990; Cutler et al., 1991; Chopra et al., 1992;

Jegadeesh and Titman, 1993, 1995; Shleifer, 2000; Lo, 2019). At the empirical level, recent works have highlighted the existence of forecasting skills related to cognitive ability and pattern recognition (Corgnet et al., 2022) and personality traits such as openness to experience (Tetlock and Gardner, 2016).[2] Furthermore, a cursory look at financial media indicates that experts and day traders continue to forecast stock markets, thus suggesting it might be a potentially profitable activity (Antweiler and Frank, 2004; Engelberg et al., 2012; Sprenger et al., 2014) or else a waste of time and money.

The early empirical analysis using data from the Livingstone Survey should also be interpreted with care. The survey collects forecasts that are not incentivized for their accuracy, thus possibly catering to other motives. For example, Pierdzioch and Rülke (2012) have shown that the Livingston experts tend to make forecasts that are markedly different from the average forecast, thus showing anti-herding behavior. This phenomenon relates to the reputational gains a forecaster can obtain from being right when all others are wrong, the so-called 'superstar effect' (Scharfstein and Stein, 1990; Laster et al., 1999). Furthermore, the Livingston survey is based only on a handful of experts, which may not represent the population of experts.[3] In contrast, the longitudinal survey employed by Ben-David et al. (2013) and Boutros et al. (2020) collected forecast data of several thousand chief financial officers.[4]

2. Our study

[2] Tetlock and Kahneman (see p. 267 of Kahneman et al., 2021) who were prominent supporters of the absence of financial forecasting expertise have revised their earlier beliefs to acknowledge the existence of superforecasters who consistently outperform their peers as well as computational models.

[3] In contrast to the finding of Pierdzioch and Rülke (2012), Graham (1999) reports that analysts with high reputation are more likely to herd based on the data from analysts who publish investment newsletters.

[4] For each quarter over a period of 10 years, 2,000 to 3,000 financial officers were sampled to answer forecasting questions (including predicting the S&P 500 over the next 10 years). The response rate varied between 5% and 8% across quarters.

Our goal is to investigate whether financial experts can forecast any better than laypeople using experiments instead of surveys. Our design is inspired by previous works using forecasting tasks in the lab to study financial forecasting by non-experts. This research highlights people's tendency to extrapolate past trends in prices (De Bondt, 1993; Bloomfield and Hales, 2002; Schwaiger et al., 2020) and their failure to anticipate mean reversion (Beshears et al., 2013). To assess forecasting expertise, we conducted our experiments with financial professionals and compared their performance with students typically used as a convenience sample in finance experiments (Kirchler et al., 2018).

Financial professionals might perform better in a setting that more closely resembles their daily working environment. Thus, it seems vital to study financial forecasting expertise not only in lab but also in field settings. Because expertise requires advanced knowledge in a specific domain (Chi, 2006), we can only expect financial experts to exhibit superior performance on tasks that closely relate to their daily work practices. In our field task, participants had to make a prediction of an actual financial series, the Nikkei index, that they regularly track and predict over the course of their workweek. They would likely have specific knowledge about how this series responds to news impacting the Japanese economy such as changes in monetary policies, inflation, GDP growth, trade and employment data. We thus *hypothesize* that finance professionals will outperform students in the field task but not in the lab tasks.

To test our *hypothesis*, we recruited certified financial analysts from Japan (Certified Member Analysts of the Securities Analysts Association of Japan, CFAs henceforth) and students from an elite Japanese university. All participants were informed that the study consisted of 4 experiments. Each experiment involved forecasting tasks that varied the type of financial series to be predicted and the information available for the forecast. Forecasting tasks were selected purposely to capture various skills ranging from basic knowledge of probability and statistics to data visualization, all of which are part of the CFA curriculum.

The three out of four experiments consisted of predicting the next value of *artificially-created* series, as in Bloomfield and Hales (2002), or *real* historical stock market series, as in Glaser et al. (2019). In the first experiment (RW, henceforth), participants had to predict random walk series. In that case, the best strategy was to recognize that the series were random and that there was no predictable pattern. In that task, we assess a forecaster's ability to identify the lack of forecastability of a series thus not exhibiting common behavioral biases. In the second experiment (SPmonthly, henceforth), participants were shown either graphs of 12 successive end-of-month prices or returns of randomly selected stocks from the S&P 500 starting from a randomly selected day between January 1st, 2008, and June 30th, 2018. However, the series to be predicted were unlabeled and graphically shown using unusual monthly frequencies. In the third experiment (SPdaily, henceforth), participants had to forecast historical stock market series, whereby participants were shown graphs of 12 months of end-of-day prices instead of 12 months of end-of-month prices. The information display was similar to what analysts would see on their computer terminals and, thus, more natural to professionals.[5] However, like the previous task, they were not told the name of the stock or the starting date of the graph, which was randomly selected.

Finally, the fourth experiment (NKI, henceforth) asked participants to make a three to

[5] On most financial websites, financial series are shown at daily frequencies, except for intraday and very long historical data. Importantly, none of these series show as few data points (12) as in the SPmonthly data. It follows that the display in SPmonthly is at odds with how financial series are typically represented.

four weeks ahead forecast on the closing price of the Nikkei.[6] This task is very similar to our finance professional subjects' daily jobs. In this task, professionals can use their financial knowledge regarding how the Nikkei typically reacts to fundamental news about the Japanese economy. This task involves an information-intensive environment, which starkly differs from the three other tasks.

The first three experiments involved lab tasks, whereas the last one featured a field task closely related to finance professionals' day-to-day jobs. In NKI, we deploy a longitudinal experiment that consists of forecasting actual financial series over a long period of time. This task is inspired by the work of Glaser et al., (2019). They asked financial professionals recruited from the ZEW Financial Market Survey to forecast the German stock market one month ahead between September 2012 and June 2015. However, this study was not incentivized and did not compare professionals' and students' performance. Our main contribution is thus to assess financial forecasting expertise by comparing the performance of professionals and students in incentivized experiments using both lab and field tasks. To our knowledge, ours is the first study comparing professionals and students across lab and field tasks.

To further explain our distinction between the three lab prediction tasks and the field prediction tasks, we refer to the classification in Harrison and List (2004). The field prediction task we implement is distinct from the lab tasks because it can be classified as a framed field experiment, which is characterized by a field context in the task and the information subjects can use (Harrison and List, 2004, p.1014). Our field task satisfies this definition as it allows traders to use any information available in the field

[6] We told participants in advance that they will be further contacted to complete a Nikkei prediction task. This was done to facilitate the recruiting of professionals as they could more easily relate to this task than to more abstract lab forecasting tasks.

to predict real-time financial series for a typical horizon of one month. In the lab tasks, participants cannot use any relevant financial information as they do not know the series, which are abstract and unlabelled. In this paper, we use the terms 'forecasts' and 'predictions' broadly, applying it to the prediction of random variables (RW) and actual financial time series (NKI). These prediction tasks vary in terms of predictability and financial knowledge used to forecast the series.

Although students and professionals performed similarly in the most abstract lab tasks that only required basic numeracy skills, their performance differed in the more realistic lab task and the field task. In the case of SPdaily, in which professionals had to forecast a historical series without any information but the graphs of the series, we found that students performed better than professionals. Some might argue that the lack of a consistent pattern in the results simply reflects the random nature of performance in the lab tasks.

However, we show that these tasks require cognitive skills and financial literacy. This is the reason why we recruited elite students who are likely to match professionals in terms of cognitive ability (Corgnet et al., 2018, 2022) while being less financially literate. To clearly identify the relative importance of cognitive ability and financial literacy in task performance, we also recruited non-elite students who possess both lower cognitive ability and lower financial literacy than professionals (see Sections 5.2.2 and 5.2.3).

Overall, our *hypothesis* that CFAs will only be able to outperform elite students in the field task is supported by our data. Combining the performance measures on the three lab tasks, we do not observe significant differences between elite students and CFAs, whereas professionals outperformed students in the field task.

Furthermore, differences in performance across samples were largely mediated by financial literacy, whereas cognitive ability did not play a substantial role. This finding suggests that CFAs possess financial forecasting expertise that is not entirely captured by cognitive skills. CFAs could outperform elite students in the field task (NKI) because they had access to additional tools, including paid services providing real-time updates on the stock market and technical analysis indicators. It follows that it might be difficult to assess the true extent of financial professionals' expertise in lab studies alone.

We discuss the generalizability of our findings following the SANS (Selection, Attrition, Naturalness, Scaling) conditions in List (2021).[7] Regarding *selection*, we were able to invite certified members of the Securities Analysts Association of Japan (more than 16,000 people) who all work in the banking and finance industry. The members of this association constitute our target population, and 1.6% of the invited members were registered in our study. These professionals had an average of more than 10 years of experience and an annual gross income that was on par with financial analysts in Japan.[8]

Out of those who registered, a large proportion participated in the first task (83.1%), and *attrition* was low for the three consecutive tasks as 93.4%, 82.1%, and 72.2% of those who participated in RW participated in SPdaily, SPmonthly, and NKI.

Regarding the *naturalness* of the setting, this is a dimension that we exogenously varied in our study so that the last forecasting task was familiar to CFAs. Indeed, a great

---

[7] We do not discuss scaling as our study does not directly suggest a large-scale policy intervention.
[8] CFAs reported an average of more than 10 years of experience and an annual gross income close to $100,000, which is slightly higher than what Japanese financial analysts typically earn (¥8,490,000 ($82,500).
(see http://www.salaryexplorer.com/salary-survey.php?loc=107&loctype=1&job=1013&jobtype=3).

majority of CFAs (74.0%) reported that they had been forecasting the Nikkei index at least once in the previous year. Despite the naturalness of the task, the stakes involved in CFAs' professional forecasting activities are likely to be larger than in our study. We purposely abstracted away from reputation issues, organizational hierarchies, and implicit incentives for promotion to ensure differences in forecasting performance between CFAs and students were only due to skills and expertise.

3. Literature review (professionals vs students)

A growing body of experimental studies has directly compared financial professionals with students or the general working population, often showing that students perform at least as well as professionals. Glaser et al. (2007) reported no performance differences in a simple task consisting of identifying a stochastic variable's trend. In a herding experiment, Cipriani and Guarino (2009) reported results for financial professionals that were very similar to those obtained with students (Cipriani and Guarino, 2005). Comparing their findings with financial professionals to previous results with students, Mann and Lock (2005) showed that both samples exhibited a disposition effect, as participants were more likely to continue to hold stocks that had lost value than stocks that had gained value. In Weitzel et al. (2020), professionals and students achieved similar earnings in experimental markets in which both pools participated in the same sessions.[9] Schwaiger et al. (2020) also found that both professionals and students exhibited more optimistic price expectations if the price of an asset fell before recovering than if the price moved in the opposite scenario. Holmen et al. (2023) compared finance professionals with the general working population. They found that finance professionals demonstrate less risk aversion, trustworthiness, and

[9] This earning comparison is not reported in the paper, but the data is publicly available and one can calculate that the difference in earnings in mixed sessions was close to zero (0.0658%) (Wilcoxon Rank Sum Test, p-value = 0.910).

greater competitiveness than their non-finance peers.

Among those studies that found differences across samples, some report professionals exhibiting stronger behavioral biases than students while others report the opposite. Among the former, Haigh and List (2005) showed that Chicago Board of Trade traders exhibited higher levels of myopic loss aversion than students. Gilad and Kliger (2008) also reported that financial professionals were more affected by priming in a laboratory investment task than students. Professionals were more likely to invest in stock when they were asked to remember a story stimulating risk-seeking behavior before making their investment decision. Relatedly, Cohn et al. (2015) showed that priming financial professionals with a boom or a bust scenario affected their risk attitudes in line with countercyclical risk aversion. Furthermore, Glaser et al. (2007) reported higher levels of overconfidence in professionals than students in a laboratory prediction task.

Among the latter, Schwaiger et al. (2020) showed that students were more likely to be impacted by the framing of the graphical information associated with the history of stock prices than professionals. In a series of herding experiments, Alevy et al. (2007) showed that finance professionals did better than students in assessing the precision of signals and were also less likely to be impacted by losses. However, the authors did not report significant differences in earnings between financial professionals and students (see Result 1, p. 161).

Interestingly, differences among financial professionals have also been identified. In particular, professionals with more extensive market experience tend to exhibit fewer biases associated with the endowment effect in various marketplaces (see Genesove and Mayer, 2001; List, 2003; Locke and Mann, 2005).

Overall, the evidence that financial professionals perform better than students in experimental tasks is somewhat limited (see Fréchette, 2015, for a review). Yet, this lack of sharp differences across samples might be due to the abstract nature of the task participants faced in the experimental studies. Previous studies have focused on the 'representativeness of the population', while abstracting away from the 'representativeness of the situation'. However, List (2021, p. VIII) claims that the 'representativeness of the situation' might be the most critical dimension. In this paper, we contribute to the literature by manipulating both the representativeness of the population, comparing financial professionals and students, and the representativeness of the task, using abstract lab tasks and a field forecasting task.

4. Design

4.1. Participants

Our participants consisted of professionals who were CFAs and students from Osaka University. Professionals were certified members of the Securities Analysts Association of Japan (SAAJ), recruited with the association's support. Out of more than 16,000 members of SAAJ, whom we invited to participate in our experiment, 255 of them initially registered for the experiment. Tasks were completed sequentially in the following order: RW, SPdaily, SPmonthly, and NKI. For SPdaily, we recruited people who completed RW. For SPmonthly, we recruited those who completed SPdaily. Although we encouraged participation in all the tasks in our e-mail invitations, it was not mandatory. RW, SPmonthly, and SPdaily were all conducted between December 2020 and March 2021. The NKI was conducted between March and September 2021. In the end, 212 CFAs participated in RW. Among them, 198 and 174 participated in SPdaily and SPmonthly, respectively. Finally, 153 participated in NKI, which was open to all CFAs who initially registered (255). The experiments were conducted online using a platform we developed on Qualtrics.

CFA is the highest-level professional certification in the financial industry. To become a CFA, candidates must have at least 3 years of professional experience in investment-related jobs. Candidates need to pass two exams with an overall success rate of between 20% and 25% on the two tests. Among Japanese CFAs, a large proportion of applicants work for the most renowned financial institutions in the country, such as Nomura Securities and Mitsubishi UFJ. Given the difficulty of the exam, they are very likely to have advanced financial knowledge, as is illustrated by the financial literacy scores of our CFA sample (91.06% of correct answers on the 12-item financial literacy scale, Fernandes et al., 2014). To the extent that we are interested in assessing how financial knowledge and experience can explain differences in performance between financial professionals and university students, CFAs represent an ideal pool of professional participants.

Our university participants were enrolled at Osaka University, widely considered an elite university in Japan. According to entry exam scores, students admitted to Osaka University are among the top 10% in Japan. According to international rankings such as QS (https://www.topuniversities.com/university-rankings/world-university-rankings/2022) and Shanghai (https://www.shanghairanking.com/), Osaka University is among the Top 5 universities in Japan. Not surprisingly, the scores on the CRT test (Frederick, 2005) of our Osaka University students' sample place them at the top of the worldwide distribution of cognitive ability along with institutions such as MIT and Princeton. Our participants were enrolled in a variety of schools including business and economics (10.1%), medical and pharmaceutical science (18.4%), engineering and technology (37.7%), humanities (23.2%) and social sciences (8.3%).[10] In RW, 228 students participated. Among them, 221 and 206 participated in SPdaily and SPmonthly. Finally, 155 participated in NKI. We note that the gender composition and average CRT scores did not differ significantly across tasks for CFAs and students (see Table C1 in Internet Appendix C). We recruited a number of participants to ensure we can detect small to medium effect sizes at a 5% level.[11]

4.2. Protocol

Participants were involved in a longitudinal study that consisted of four forecasting tasks.

a) **RW**: **Forecasting random walk time series**. We asked participants to predict the pre-generated (artificial) time series that follow a random walk.
b) **SPmonthly: Forecasting S&P 500 stocks prices or stock returns based on monthly series**. We asked participants to predict either stock prices or stock returns from the S&P 500 index without knowing the actual name of the stock and the selected historical sample. Participants received graphical information of monthly stock values shown as prices or returns.

[10] There were 5 participants without information about their school affiliation. We recruited broadly across schools to ensure our student participants possessed high cognitive skills while not being specifically trained in finance. Our aim was to isolate the financial expertise of CFAs, which depends both on their training and experience.

[11] For our power calculations, we started with RW, which was the first one conducted with our participants. We determined that in order to identify an effect size of 0.25 of standard deviation at a 5% significance we needed about 250 observations per sample. We ended up with 212 and 228 observations for the sample of CFAs and elite students, which was slightly lower than expected. Yet, we obtained enough power to detect low to moderate effect sizes. This was the main power calculation we used. Regarding the NKI, any calculation of power would surely depend on attrition. In order to identify an effect size of 0.25 of standard deviation at a 5% significance level across samples we calculated that we also needed about 250 observations per sample, which would be achieved even under extremely high attrition levels above 80%. We ended up with 717 and 538 observations for the sample of CFAs and elite students.

c) **SPdaily**: **Forecasting S&P 500 stocks prices based on daily series**. This task was similar to the SPmonthly task, except that participants received information about the daily instead of the monthly stock values.

d) **NKI: Forecasting the Nikkei index**. We asked participants to forecast the closing price of the Nikkei index three to four weeks ahead.

Participants knew they were recruited for four tasks and were told that each of the first three tasks would take between 15 and 30 minutes. While participants were not given instructions for each task when recruited, they were told that the last task consisted of forecasting the Nikkei index. As noted above, each task was completed sequentially. For SPdaily, we recruited people who completed RW. For SPmonthly, we recruited those who completed SPdaily.[12] Although we encouraged participation in all the tasks in our e-mail invitations, it was not mandatory. Table 1 summarizes the main features of the four forecasting tasks.

[TABLE 1 AROUND HERE]

For each of the first three tasks, CFAs (students) were paid ¥1,000 (¥200) for their participation, which is about $9.7 ($1.94) at the start of the study.[13] They also had a 10% chance of being selected for an additional reward based on their earned points in the experiment, where 1 point was worth ¥100 ($0.97) for CFAs and ¥20 ($0.19) for students. The payment scheme for NKI was based on a tournament structure where the best CFA (student) performer earned ¥5,000 (¥1,000) [$48.5 ($9.7)]. Kirchler et al. (2018) reported no differences between financial professionals and students in their

[12] Professional participants were invited to RW, SPdaily, and SPmonthly (in this order) once a week. In particular, those who had completed RW were subsequently informed of their payment and then recruited for SPdaily and SPmonthly. For students, RW, SPdaily, and SPmonthly were conducted two days apart.
[13] We use the average exchange rate for the month in which the study started (December, 2020) whenever converting yens to dollars.

responses to tournament incentives. We thus expect that the use of a tournament scheme in NKI will not alter the comparison of students and professionals.[14]

In all tasks, the payments offered to CFAs were five times greater than for students to compensate for the difference in hourly wages for these two populations, considering the hourly wage of undergraduate students at the university is about ¥950 ($9.65), and the average hourly wage of CFAs in Japan is about ¥4,300 ($41.71).[15] All the payments were made using Amazon gift cards (e-mail version), and the mode of payment was known to participants when signing up for the experiment. Below, we describe each task in detail.

4.3. Tasks

4.3.1. RW: Predicting Random Walk Time Series

Our first task aimed to capture basic quantitative skills related to understanding probabilities and randomness. Such skills are also part of the CFA Level I curriculum (see Section II.B in the candidate book of knowledge).[16] The task was based on the experimental design of Bloomfield and Hales (2002).[17] Participants were shown 16 graphs that were generated using a random walk stochastic process following Bloomfield and Hales (2002, p. 403), and they were told that:

---

[14] The authors reported a difference across samples when using rank incentives, which we do not use in our study for that reason.

[15] According to information available on a Japanese website related to change of occupation (https://mynavi-agent.jp/helpful/income2020/category/finance01/, last checked on Oct 28, 2021), the average annual income of financial analysts is ¥8,240,000 ($79,928). With standard working hours (8 hours/day, 20 days/months), it is approximately ¥4,300 ($41.71) per hour.

[16] https://www.cfainstitute.org/en/programs/cfa/curriculum/cbok/cbok-2022

[17] Asparouhova et al. (2009) rightly noted that the sequences shown to participants in Bloomfield and Hales (2002) are not representative of what would be observed under a random walk and this might have induced participants to exhibit regime-switching beliefs. However, this issue is less of a concern in our study because, unlike Bloomfield and Hales (2002), we are not testing the regime-switching model of Barberis et al. (1998) but simply using their experimental design as a forecasting task.

"We have constructed a model of a random process that works much like flipping a fair coin. Using this model, we have created sequences of outcomes. An upward movement indicates a "heads" outcome, and a downward movement indicates a "tails" outcome."

Each graph consisted of nine prices and captured both upward and downward sloping dynamics (see Internet Appendix A.1 for graphs and detailed instructions).[18]

For each graph, we gave participants one unit of an asset whose value could be either 0 or 100 points depending on the next movement of the graph (0 if it went down and 100 if it went up). After observing each graph, participants stated a price (an integer between 0 and 100) at which they were willing to sell the asset back to the experimenter. After participants submitted their prices, one of the 16 series was randomly selected for payment. We used a Becker et al. (1964) mechanism (BDM, henceforth) to elicit participants' beliefs about the next movement of the random walk. That is, the price stated by the participant was compared to a randomly generated integer from the interval [0,100]. If the random number was less than or equal to the stated price, participants kept the asset, and the payoff was determined by the subsequent movement of the price chart. If the random number was strictly greater than the stated price, participants were paid the random number regardless of the next movement of the price chart.[19]

---

[18] The English translation of the instructions can also be found at: https://bgt.au1.qualtrics.com/jfe/form/SV_9RAxWjjGFww0zSl.

[19] This incentivization procedure differs from Bloomfield and Hales (2002). In Bloomfield and Hales (2002), participants were told that if they stated a price above 50, they were buying one unit of the asset at each (integer) price between 51 and the stated price, while if they stated a price below 50, they were selling one unit of the asset at each (integer) price between the stated price and 49. For example, if a participant stated a price of 54, he or she bought 4 units of the asset at the following prices: 51, 52, 53 and 54. If the next movement of the random walk was up, then payments were equal to 400 - (51 + 52 + 53 + 54) = 190. We chose to implement a standard BDM mechanism for the sake of simplifying the instructions and facilitating the understanding of the task.

212 professionals (CFAs) and 228 students participated in the task, lasting around 20 minutes.[20] These professionals had an average of more than 10 years of experience and an annual gross income greater than ¥10M ($97,000). CFAs earned an average of ¥1,362 ($13.21), and the maximum payment was ¥11,000 ($106.7).[21] Students earned an average of ¥316 ($3.65), and the maximum payment was ¥3,160 ($30.65).[22]

In addition, we elicited participants' risk attitudes (using a variant of the multiple price list of Holt and Laury, 2002), loss attitudes (using the task employed by Kirchler et al. 2018), cognitive reflection test scores (CRT, Frederick, 2005), and basic demographics (see Internet Appendix B for details on these additional measures). We use CRT as a proxy of cognitive ability as it has been shown to strongly correlate with conventional measures of fluid intelligence (see Frederick, 2005; Corgnet et al., 2018, 2022). We also administered a financial literacy quiz (Fernandes et al., 2014).

4.3.2. SPmonthly: Forecasting S&P 500 stocks prices or stock returns based on monthly series

Our second task aimed to capture data visualization skills, which are also part of the CFA Level I curriculum (see Section II.C in the candidate book of knowledge).[23] The CFA module of data visualization includes understanding how returns are calculated and graphically represented. We used a forecasting task similar to the one in Glaser et al. (2019) that asks participants to forecast either returns or prices of financial series. In line with their study, we conducted a 2×2 between-subject factorial design giving rise to a total of 4 treatments, where the first dimension varied

[20] We excluded those who accessed from the same IP address or who answered multiple times with the same participant ID. This led us to drop 13 CFA responses. No responses were discarded for students.
[21] The average payment for those who have been selected for the additional reward was ¥6,120 ($59.36).
[22] The average payment for those who have been selected for the additional reward was ¥1,524 ($14.78).
[23] https://www.cfainstitute.org/en/programs/cfa/curriculum/cbok/cbok-2022

the *type of information* shown on the graph, either *price* or *return*, and the second dimension varied the *forecasting variable*, either *price* or *return*. In all treatments, we added the final price of the corresponding series to the chart. Unlike Glaser et al. (2019), we conducted all four treatments with the same series for both professionals and students.

As in Glaser et al. (2019), we simulated financial series using actual stock market data. In particular, we used historical S&P 500 stock end-of-day prices between January 1st, 2008, and June 30th, 2018. We selected a day at random and then constructed a price chart of 12 data points corresponding to end-of-month prices over a year. The end-of-month prices were rescaled so that starting prices were equal to 100. Participants were neither told the name of the stock nor the corresponding time period that was randomly selected. Given our procedure, there were 1,916,250 possible charts created so that participants could not practically identify the financial series. It follows that participants could only use the chart, i.e., the apparent price patterns, to make predictions. Furthermore, this procedure was such that the generated line charts did not resemble the graphical information professionals typically face in their daily job (see Internet Appendix A.2 for graphs and detailed instructions).[24-25] We refer to these charts as *unnatural-display* series in Table 1.

In the two treatments in which participants had to forecast prices, we asked them to predict the price in the next period given the corresponding chart. Performance (in points) on the task was measured by:

[24] The English translation of the instructions can also be found at: https://bgt.au1.qualtrics.com/jfe/form/SV_74H7IiRiVqUQCzA.
[25] Glaser et al. (2019) used bar charts for returns and line charts for prices because they are the standard way of showing this information. We have opted for using line charts for both variables to limit the differences between the two treatments.

$$max\ \{200 - 1000 \times \left|\frac{price\ forecast - realized\ price}{realized\ price}\right|, 0\}$$

For the other two treatments in which participants had to forecast returns, we asked them to predict the next return given the corresponding chart. Performance (in points) on the task was measured by:

$$max\ \{200 - 10 \times |return\ forecast - realized\ return\ |, 0\}$$

Where:

$$return = \frac{realized\ price - last\ price}{last\ price} \times 100$$

Each participant was given 20 series sequentially, thus producing 20 forecasts, regardless of the treatment. The order of the series was randomized across participants.

In total, 174 professionals (CFAs) and 206 students participated in this task, which lasted about 20 minutes.[26] Each professional (student) was assigned at random to one of the four treatments. CFAs earned an average of ¥1,675 ($16.25), and the maximum payment was ¥10,930 ($106.02).[27] Students earned an average of ¥284 ($2.75), and the maximum payment was ¥2,170 ($21.05).[28]

We referred to the treatments wherein the information provided on the chart (price or return) and the variable to be forecasted (price or return) are the same as ***congruent*** treatments. In contrast, we referred to the treatments in which the

[26] We excluded those who accessed from the same IP address or who accessed several times with the same participant ID. As a result, 25 responses from CFAs and 0 response from students were discarded.
[27] The average payment for those who have been selected for the additional reward was ¥8,338 ($80.88).
[28] The average payment for those who have been selected for the additional reward was ¥1,441 ($13.98).

information provided on the chart differed from the variable to be forecasted as ***incongruent*** treatments.

4.3.3. SPdaily: **Forecasting S&P 500 stocks prices based on daily series**.

Our third task, which captures the skills directly related to the use of charts for forecasting, is part of the CFA Level I curriculum (see Section X.G in the candidate book of knowledge on technical analysis).[29] We asked participants to make 20 price forecasts of the historical S&P 500 series in this task. Before making a forecast, participants were presented with a chart showing the end-of-day prices of an S&P 500 stock during a year. The graph was created based on the price of randomly selected stocks from the S&P 500 starting from a randomly selected day between January 1st, 2008, and June 30th, 2018. The end-of-day prices were rescaled so that starting prices were equal to 100. Participants were neither told the name of the stock nor the corresponding time period that was randomly selected. This task is similar to the SPmonthly task, except that participants received information about the daily instead of the monthly stock price movements. Unlike SPmonthly, the series used in SPdaily resembled the graphical information professionals would typically face (see Internet Appendix A.3 for graphs and detailed instructions), as is shown in Figure 1.[30] We refer to these charts as *natural-display* series in Table 1.

[FIGURE 1 AROUND HERE]

For each of the 20 charts, participants were asked to forecast the end-of-day price of the stock 30 days after the last price displayed on the graph. All participants

[29] https://www.cfainstitute.org/en/programs/cfa/curriculum/cbok/cbok-2022

[30] The English translation of the instructions can also be found at: https://bgt.au1.qualtrics.com/jfe/form/SV_0pmGV8LB9DMRPc9.

made price forecasts based on the same 20 time-series, but the order in which these graphs were shown was randomized across participants.

The payoff for this task depended on the forecast accuracy for one randomly selected series. Similar to SPmonthly, the number of points was calculated as follows:

$$max\left\{200 - 1000 \times \left|\frac{price\ forecast - realized\ price}{realized\ price}\right|, 0\right\}$$

In total, 198 professionals (CFAs) and 221 students participated in this task, which lasted about 20 minutes.[31] CFAs earned an average of ¥1,840 ($17.85), and the maximum payment was ¥10,980 ($106.51).[32] Students earned an average of ¥315 ($3.06), and the maximum payment was ¥2,120 ($20.56).[33]

4.3.4. NKI: Forecasting the Nikkei index

Our fourth task, which captures skills related to actual market forecasts, is part of the CFA Level III curriculum (see Section III.J in the candidate book of knowledge on capital market expectations). For this task, forecasters can use technical analysis similarly to the previous task.[34] The use of technical analysis is easier than in SPdaily because forecasters will be able to use standard technical tools available on the Nikkei series.[35] This task also requires broader knowledge of the functioning of financial markets. Financial professionals would, for example, know better the factors that have impacts on stock markets such as key economic indicators and international politics.

[31] We excluded 6 CFA responses because of multiple entries from the same IP address or using the same participant ID. There was no response excluded for students.
[32] The average payment for those who were selected for the additional reward was ¥8,233 ($79.86).
[33] The average payment for those who were selected for the additional reward was ¥1,694 ($16.43).
[34] For empirical evidence, see Menkhoff (2010).
[35] https://www.cfainstitute.org/en/programs/cfa/curriculum/cbok/cbok-2022

[TABLE 2 AROUND HERE]

In this task, we asked participants to predict the Nikkei index once a month for several months.[36-37] The forecasting horizon for the Nikkei was defined after consulting members of SAAJ to ensure it was consistent with the type of forecasting exercises CFAs practice on a regular basis.

Unlike SPdaily, which relies on historical data, the NKI uses '*real-time*' data. Participants had to forecast the closing price of the Nikkei index three to four weeks after a given deadline. As is shown in Table 2, participants had 3 or 4 days to submit their forecasts (forecast window) for the closing value of the Nikkei. During these 3 or 4 days, they can freely use all sorts of resources to help enhance the quality of their decisions, e.g., they can visit professional websites or use professional tools. The number of participants varied across waves, as shown in the last column of Table 2.

While the payoff in this task depended on the forecasting accuracy, unlike SPmonthly and SPdaily, a tournament incentive was used. In particular, the most accurate CFA or student forecaster, i.e., the one whose forecast was closest to the closing value of the Nikkei, for a given forecasting day each session received an Amazon gift card of ¥5,000 (¥1,000) [$48.5 ($9.7)], there was no participation fee for this task. In case of a tie, one winner was chosen randomly among the top performers (see Internet Appendix A.4 for detailed instructions).[38-39]

[36] As for the previous tasks, we excluded those who accessed from the same IP address or who accessed several times with the same participant ID.
[37] We also asked professionals to predict the Nikkei in February (16-19), but we do not report the results here because we did not elicit students' forecasts.
[38] Ties happened in Wave 1 and 3 for students (2 participants) and in Wave 3 for CFAs (3 participants).
[39] The English translation of the instructions can also be found at: https://bgt.au1.qualtrics.com/jfe/form/SV_1MNumMb42kKnpDo

After each forecasting day, participants received graphical feedback, separately for students and CFAs, showing the realized closing Nikkei index and their forecasts and the average forecast. The screen also showed the most accurate and the second and third most accurate forecasts.

In the June forecast period, we added a questionnaire to the online platform to assess participants' prior experience in forecasting the Nikkei index, the effort they exerted on the forecasting task, and the use of professional services (see Internet Appendix A.4.2).

5. Results

5.1. Forecasting performance of CFAs and students

We compare the forecasts of financial professionals and students by using a measure of relative forecasting error (RFE), which is defined as follows:

$$100 \times \frac{|forecast - value|}{value}.$$

Where *value* is the actual value of the series to be forecasted for SPmonthly, SPdaily, and NKI, and the bid value that maximizes forecasting performance in RW (i.e., 50).[40] We use the RFE measure because participants were incentivized to minimize forecasting errors in all tasks. In Figure 2, we show the standardized values of RFE calculated at the participant's level.

[FIGURE 2 AROUND HERE]

[TABLE 3 AROUND HERE]

[40] Similarly, Corgnet et al., (2022) use *price forecasting error* as their accuracy measure, which is defined as RFE × value.

In Table 3, we show that CFAs and students performed similarly in RW (see regression (1)), which is not surprising because this task did not require specific financial skills. The independent variable is the standardized *RFE*[41]*,* the dependent variables are *CFA Dummy* takes value one for participants who are CFAs. A higher *CRT score* indicates higher cognitive skills. The *Risk Aversion* is measured by the number of safe options. Yet, participants who scored higher on cognitive ability (as measured with CRT) performed better in RW. This result is in line with previous research showing that people with high CRT scores tend to better understand the concepts of probability and randomness (see, e.g., Toplak et al., 2011, 2014).

Although we used RFE as a common measure of forecasting performance across all tasks, it could be impacted by risk attitudes in RW. For example, risk-averse participants will tend to bid below the expected value of the asset, which is 50. In regression (1') of Table 3, we thus added risk attitudes as a control variable in our regressions. The CFA dummy remains insignificant when adding risk attitudes as a control variable. In addition, risk attitude does not appear to explain forecasting performance significantly. Finally, we do not find significant differences in the risk attitudes of CFAs and students (Wilcoxon Rank Sum Test, p-value = 0.160, see Table 5) although, as expected, CFAs were slightly less risk-averse, a result which is generally consistent with the result of Holmen et.al (2023).

The SPmonthly task requires financial skills related to data visualization and returns numeracy (CFA exam, Section II.C) (Glaser et al., 2019) that are especially relevant to financial professionals and a key component of financial literacy (see, e.g.,

[41] That is, transforming RFE so that is has a mean of 0 and a standard deviation of 1.

Fernandes et al., 2014). Thus, it is not surprising that CFAs outperformed students in SPmonthly (see regression (2) in Table 3).

[TABLE 4 AROUND HERE]

In Table 4, we show that the higher performance of CFAs in SPmonthly is due to the incongruent treatments in which the variable to be predicted (either returns or prices) differed from the variable that was shown on the chart (either returns or prices) (see regression (1)). The definition of variables is the same as in Table 3. In that case, CFAs might have performed better than students because they possess greater numeracy skills (see comparison of financial literacy scores across samples in Table 5), which are required to convert prices to returns and vice versa.

For the two *congruent* treatments in which the variable to be predicted and the chart information coincided, numeracy skills were arguably less important because participants no longer had to make calculations to convert prices to returns and vice versa. This might explain why we observed no differences between CFAs and students in that case (see regression (2) in Table 4).

Participants forecasted actual financial series in SPdaily. However, this task was performed in a controlled setting in which financial professionals could not identify the series and did not have access to news feeds and their usual analytical tools. This context implies that CFAs could not outperform students even though CFAs, unlike students, have access to a range of professional tools and have greater technical knowledge. Interestingly, CFAs underperformed students (see regression (3) in Table 3). In contrast, CFAs outperformed students (see regression (4) in Table 3) in the field task (NKI) in which they forecasted financial series in a natural longitudinal context in which they knew the series (Nikkei), and they could employ their usual professional tools and knowledge.[42] Interestingly, CRT led to lower relative forecasting errors in RW, SPmonthly, and SPdaily (and significantly so for RW) whereas it was not the case for NKI. This suggests that performing well in our field task required expertise that could not solely be captured by standard cognitive skills.

We use four tasks, thus completing four comparison tests between CFAs and students. One might argue that any identified differences are due to multiple hypotheses testing. However, the likelihood of having at least 3 out of 4 independent tests being significant at the 5% level is less than 0.05%. Yet, one could argue that forecasting performance across tasks is highly correlated. To calculate the correlation of RFE across tasks, we compute the average RFE for each participant in each task. We then calculate the six pairwise correlation coefficients across the four tasks. None of these correlations are significantly different from zero, and two are negative. The average value of the six pairwise correlation coefficients is equal to 0.013. Standard multiple hypotheses corrections (à la Holm, 1979) give us the following *p*-values when comparing CFAs and students for SPmonthly, SPdaily, and NKI: 0.078, 0.080, and 0.080. The fact that our outcome measures (RFE) are not significantly correlated suggests similar results would be obtained using more recent multiple hypotheses testing techniques that account for dependence across measures (List et al., 2019).[43]

[TABLE 5 AROUND HERE]

[42] The median and modal response for students was that they never predicted the Nikkei index whereas CFAs median and modal response was that they predict the Nikkei index once a month.
[43] We are not aware of multiple hypothesis techniques that could provide *p*-values corrections that account for dependence across measures while allowing for the estimation of panel regressions. We thus proceeded differently as explained in Table C2.

We should also note that our *hypothesis* only implied two tests. The first test compares the performance of CFAs and students in the lab tasks, while the second test assesses their relative performance in the field task.

5.2. Financial literacy, cognitive skills, and forecasting performance

This section aims to uncover the mechanisms by which CFAs' forecasting performance differs from students' performance by inquiring on the mediating role of cognitive ability and financial literacy.

5.2.1. Mediating role of CRT and financial literacy

In Tables 3 and 4, we assess the effect of being a CFA on relative forecast error while controlling for CRT scores. In Tables C4 and C5 in the Internet Appendix, we conduct the same regressions without controlling CRT to assess its role as a mediator. We do not find evidence of a mediating role of CRT as the coefficient associated with CFA is not substantially impacted by the inclusion or exclusion of CRT as a control. This is perhaps not surprising, given that there is no significant difference between the CRT scores of students and CFAs (Wilcoxon Rank Sum Test, p-value = 0.396).

Although the two samples do not seem to differ in cognitive skills, they differ in financial literacy test scores. Indeed, CFAs need to obtain very high scores on these tests to obtain their accreditation and should thus outperform even smart students. One month after Wave 6 of NKI, we collected financial literacy scores on financial professionals. We obtained 110 answers, which corresponds to 81.3% of the participants in the NKI task over all waves.[44] Not surprisingly, professionals scored

[44] Unlike students, we did not collect financial literacy for the 212 CFA participants in the RW. The reason we did not originally elicit CFAs financial literacy was that we strongly believed that they would obtain almost-perfect scores. We were also concerned about offending CFAs by asking them simple financial literacy questions, which might have the unintended consequence of discouraging them from participating in other tasks.

substantially higher on financial literacy than students (p-value < 0.001). We should note that although CRT and financial literacy scores are positively correlated, the magnitude of the correlation coefficient is moderate ($\rho = 0.324$) so that they likely cover distinct constructs.

We then assessed the mediating role of financial literacy in predicting relative forecast error using the same procedure as for CRT (see Tables C6 to C9 in the Internet Appendix).[45] We show that financial literacy plays an important role in understanding the difference in forecasting performance between CFAs and students. For the two forecasting tasks in which CFAs outperform students (the incongruent treatments in SPmonthly) and NKI, the magnitude of the coefficient associated with CFAs dummy decreases by 27.01% and 62.18% for the incongruent treatments of the SPmonthly and the NKI when we introduce financial literacy in the regression (see regression (3) in Table C9 and regression (4) in Table C8, respectively). In sum, although CFAs substantially outperformed students when controlling for CRT, differences were small when controlling for financial literacy. This result is especially remarkable given that CRT has been shown to correlate, although moderately, with numeracy skills and financial literacy (e.g., Oechssler et al., 2009; Pennycook et al., 2012; Shenhav et al., 2012; Weller et al., 2013; Campitelli and Gerrans, 2014). It follows that the expertise of professionals in financial forecasting closely relates to their knowledge of financial markets. Beyond the type of financial knowledge captured by financial literacy, expertise could also be driven by experience. However, we show in Table C8 (regression (5)) that CFAs' experience does not explain forecasting performance.

[45] We conducted the same analysis for CRT using the CFA sample of those who completed the financial literacy questionnaire to alleviate selection issues (see Tables C6, C7 and C9). It is reassuring that we obtain the same finding of the absence of a mediation role of CRT for the restricted sample as for the whole sample.

The higher performance of CFAs could also be due to greater dedication to the NKI task. Because the NKI task was likely perceived as a familiar activity, professionals might have been more engaged in the task and more attentive to market data when making their predictions. Although CFAs did wait longer (0.31 days, on average) than students to make their forecasts, possibly to observe the latest value of the Nikkei index within the forecasting window (see regression (1) in Table C10 in Internet Appendix C), controlling for the date at which a forecast was made did not alter the findings (see regression (2) in Table C10 in Internet Appendix C).

In the same vein, we controlled for the level of interest and effort participants put into the task to assess whether the superior performance of CFAs was simply due to their greater dedication to the task. To measure task dedication, we adapted the scale of Ryan (1982) with a total of 5 items such as "I put a lot of effort in the forecasting task" evaluated on a 7-point Likert scale (1- not all true to 7- very true). This effort and dedication index (Cronbach's α = 0.783) was, however, not significantly different between CFAs (3.320) and students (3.081) (Wilcoxon Rank Sum Test, p-value = 0.143). The magnitude of the coefficient for CFA Dummy only decreases by 12.1% when controlling for the dedication index (see regression (3) in Table C10 in Internet Appendix C), thus suggesting that the mediating role of dedication is limited.

CFAs could outperform elite students in the field task (NKI) because they had access to additional tools, including paid services providing real-time updates on the stock market and technical analysis indicators. We also find that the mediating role of the variables related to the frequency of paid services in the forecasting task was small (see regression (4) in Table C10) as the coefficient for CFA Dummy only decreases by 15.70% when controlling for paid services.

These results could also help discard an alternative explanation to our results that relies on the higher inherent competitiveness of finance professionals (Kirchler et al. 2018; Holmen et al. 2023). In principle, the higher level of performance by CFAs on the NKI could indeed be due to professionals being more competitive in tasks that appear relevant to their expertise and for which they are expected to excel. However, the fact that dedication to the task and the use of paid services could not fully account for the superior performance of CFAs over students in the NKI suggests that heightened competitiveness is not the primary explanation for our findings.

However, we should note that because financial literacy is only partially mediating the differences between CFAs and elite students, future research should contemplate additional moderators. Regarding the SPmonthly, this endeavor could start by measuring numeracy skills more precisely rather than using overall financial literacy scores (Cokely et al. 2012). Other relevant variables could assess the extent to which professionals use charts and which type of displays they are familiar with. Direct professional experience related to the calculation of interest rates and to the use of financial charts could also play a critical role beyond financial literacy.

5.2.2. Elite and non-elite students

The sample of students from Osaka University is characterized by high levels of cognitive skills comparable to CFAs. These students scored 88.6% of the CRT questions correctly, placing them at the top US schools' level and in the top 20% of a standard US student population (see Frederick, 2005; Corgnet et al., 2018). As a result, we will refer to Osaka University students as elite students (E-Students, henceforth).

To further study expertise and connect it with cognitive ability and financial literacy, we collected data from another sample of students from a less selective university (NE-

Students, henceforth) for the three lab tasks in September 2021.[46] In Table 5, we show that NE-Students score significantly lower than E-Students on cognitive ability and financial literacy (p-values < 0.001).[47] By contrast, E-Students scored as well as CFAs on cognitive ability while CFAs outperformed NE-Students (p-value < 0.001). As expected, CFAs outperformed all students in financial literacy. Figure 3 shows that NE-Students tend to produce larger forecasting errors than E-Students and CFAs.

[FIGURE 3 AROUND HERE]

In Table 6, we find that NE-Students performed significantly worse than E-Students and CFAs in RW and SPmonthly. Here, *NE-Student Dummy* takes value one for participants who belong to the non-elite university, while others follow the exact definition as in Table 3. In SPdaily, the performance of NE-Students and CFAs did not differ significantly. NE-Students performed worse than E-Students in SPdaily, but this difference is not statistically significant. We also note that these findings are qualitatively similar whether CRT scores are controlled for (regressions (2), (4), and (6)) or not (regressions (1), (3), and (5)). Yet, the magnitude of the NE-Student Dummy is reduced by 16.15%, 26.92%, and 74.89% for RW, SPmonthly and SPdaily when controlling for CRT scores. This result implies that a substantial part of the increase in performance of E-Students compared to NE-Students is mediated by cognitive ability. Interestingly, the share of forecasting performance mediated by cognitive ability seems to increase as the task is less abstract and more realistic.

[46] Because these experiments were conducted in September 2021, we could not gather data for the NKI for this additional sample. Our participants were enrolled in a variety of schools including science and technology (12.1%), business and economics (21.5%), social sciences (41.6%), and humanities (24.8%).
[47] In Table C11 in the Internet Appendix, we provide the same descriptive statistics for each task separately.

Table 6 focuses on cognitive ability as a mediator of the forecasting performance of NE-Students, but financial literacy, as we have seen in the previous section, might also play an important role in understanding the performance of NE-Students. The result of this analysis is shown in Table 7. To provide a meaningful comparison of the mediating role of CRT and financial literacy on the forecasting performance of NE-Students, we re-estimated the effect of CRT restricting our sample to the participants who answered both the financial literacy and CRT questions (see regressions (2), (6) and (10)). We observe that, overall, the mediating role of financial literacy is similar to the one of cognitive ability in explaining the forecasting performance of NE-Students.

[TABLE 6 AROUND HERE]

[TABLE 7 AROUND HERE]

These findings put forth that the performance in the lab tasks requires cognitive skills and cannot be attributed to chance alone.

5.2.3. Forecasting as a game of chance?

The critical role of cognitive ability as measured with CRT in explaining forecasting performance suggests that our forecasting tasks require an innate ability and are not purely a game of chance. In particular, the lab tasks we employed were originally used to highlight cognitive biases that led to poor performance (Bloomfield and Hales, 2002; Glaser et al., 2019). It is thus not surprising that higher CRT scores, which have been found to help people avoid common decision-making biases more so than standard cognitive tests (Toplak et al., 2011), predict the performance in our forecasting tasks well. It follows that elite students and professionals are more likely to avoid common decision-making biases associated with forecasting financial series than non-elite students.

These findings, however, do not challenge the efficient market hypothesis since they do not imply that markets can be consistently predicted. Indeed, if markets satisfy weak-form efficiency (Fama, 1970), then the forecast that minimizes the RFE equals the last observed value of a financial series. However, forecasts of CFAs and students deviated on average by 4.7% from the value of the last observable index values (3.9% for CFAs and 5.6% for students, Wilcoxon Rank Sum Test, p-value = 0.159). More importantly, the forecasts of both students and CFAs produced forecasting errors (RFE = 7.1% and 5.1%, respectively) that were significantly higher than those associated with using the last observable value of the index (RFE = 2.1%, Sign Rank Tests, p-values < 0.001).

In sum, our findings imply that cognitive ability and financial literacy can alleviate common biases in financial forecasting while not challenging the fact that predicting stock markets is akin to predicting a chance event. Yet, understanding chance requires skills so that predicting a chance event is not a game of chance.[48]

6. Conclusion

The forecasting skills of finance experts have often been mocked (see Kahneman, 2011) and compared to those of other primates (Tetlock, 2009a,b). Focusing on human primates only, our study showed that although financial professionals could outperform elite students on our field task, they did not consistently do so in our lab tasks. Cognitive skills, such as CRT scores, did not account for the difference in performance between professionals and students. Both groups scored similarly on cognitive ability, but differences in financial literacy between the two samples explained a significant part of the discrepancy.

Our study has three main implications. First, finance professionals do not appear to

[48] We leave it open to debate whether these skills can be considered a form of financial expertise.

consistently outperform students. Instead, in line with our hypothesis, professionals outperformed students only in the field task. Second, the extent of financial professionals' skills cannot be reliably measured using only lab tasks. Third, we showed that the difference between CFAs and students in forecasting performance was mediated by financial literacy but not by CRT scores, thus suggesting financial knowledge rather than cognitive skills explains performance in financial forecasting tasks.[49] While financial professionals outperformed students in the field task, their forecasting error was higher than that obtained by using the last observed value of the index, in line with the weak-form efficiency of markets.

Our findings provide support for the conjecture proposed in Weitzel et al. (2020) regarding the importance of financial knowledge in explaining professionals' superior performance: "*Hence, we conjecture that "professional skills" that may be rooted in real-world market experience, possibly including a more intuitive understanding of markets that goes beyond specific cognitive skills , may affect trading behavior, and lead to more efficient pricing.*" (p. 2686). We also complement this conjecture by showing that any advantage related to financial knowledge is more likely to apply for tasks that most closely resemble field tasks.

Although we have emphasized the importance of financial knowledge, instead of cognitive skills, our our results comparing elite and non-elite students are in line with previous results in cognitive finance. In particular, we show that a substantial part of the increase in performance of E-Students compared to NE-Students is mediated by cognitive ability. This result complements a rich literature using financial tasks in the lab to demonstrate the critical role of cognitive ability in explaining performance among student

[49] Though the scope of financial knowledge possessed and used by financial professionals in their daily work is different from, usually much wider than financial literacy.

participants (see Noussair et al. 2016; Corgnet et al., 2018; Hefti et al. 2018; Bosch-Rosa and Corgnet, 2022).

Our design has focused on forecasting prices in financial markets but future research could adopt a similar protocol to study accounting (e.g., Barron et al., 1998; Clement, 1999; Easterwood and Nutt, 1999; Lim, 2002; Hong and Kubik, 2003; Clement and Tse, 2005; Da and Huang, 2020), industrial (e.g., Gay et al., 2009; Pierdzioch et al., 2013; Fernandez-Perez et al., 2020) and macroeconomic forecasts (e.g., Laster et al., 1999; Carroll, 2003). Although our experiment elicited point forecasts, future research can also elicit interval and quantile forecasts, which are widely used by professionals in the retail industry (Jain et al., 2013; Gaba et al., 2017; Chen et al., 2022).

References

Abbink, K., & Rockenbach, B. (2006) "Option pricing by students and professional traders: a behavioral investigation," *Managerial and Decision Economics*, 27(6), 497-510.

Ackert, L. F., Church, B. K., & Ely, K. (2008a). Biases in individual forecasts: experimental evidence. *The Journal of Behavioral Finance*, *9*(2), 53-61.

Ackert, L. F., Church, B. K., & Shehata, M. (1997). An experimental examination of the effects of forecast bias on individuals' use of forecasted information. *Journal of Accounting Research*, *35*(1), 25-42.

Ackert, L. F., Church, B. K., & Zhang, P. (2008b). What affects the market's ability to adjust for optimistic forecast bias? Evidence from experimental asset markets. *Journal of Economic Behavior & Organization*, *66*(2), 358-372.

Alevy, J.E., Haigh, M.S., & List, J.A. (2007) "Information cascades: evidence from a field experiment with financial market professionals," *The Journal of Finance*, 62(1), 151-180.

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. Chapter 21 in Kahneman, D., Slovic, P., Tversky, A., ed. *Judgment Under Uncertainty Heuristics and Biases*, 294 -305. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511809477.022

Antweiler, W., & Frank, M. Z. (2004). “Is all that talk just noise? The information content of internet stock message boards,” *Journal of Finance*, 59(3), 1259-1294.

Asparouhova, E., Hertzel, M., & Lemmon, M. (2009). “Inference from streaks in random outcomes: Experimental evidence on beliefs in regime shifting and the law of small numbers,” *Management Science*, 55(11), 1766-1782.

Barber, B. & Odean, T. (2001) “Boys will be boys: Gender, overconfidence, and common stock investment,” *Quarterly Journal of Economics*, 116(1), 261-292.

Barberis, N., Shleifer, A., & Vishny, R. (1998). “A model of investor sentiment,” Journal of Financial Economics, 49(3), 307-343.

Barron, O. E., O. Kim, S. C. Lim, and D. E. Stevens (1998) “Using Analysts’ Forecasts to Measure Properties of Analysts’ Information Environment,” *Accounting Review* 73: 421-433.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964) “Measuring utility by a single-response sequential method,” *Behavioral Science*, 9(3), 226-232.

Ben-David, I., Graham, J. R., & Harvey, C. R. (2013) “Managerial miscalibration,” *The Quarterly Journal of Economics*, 128(4), 1547-1584.

Beshears, J., Choi, J. J., Fuster, A., Laibson, D., & Madrian, B. C. (2013) “What goes up must come down? Experimental evidence on intuitive forecasting,” *American Economic Review*, 103(3), 570-74.

Bloomfield, R., & Hales, J. (2002). Predicting the next step of a random walk: Experimental evidence of regime-shifting beliefs. *Journal of Financial Economics*, 65 (3), 397–414

Bosch-Rosa, C., & Corgnet, B. (2022). Cognitive finance. In Handbook of Experimental Finance (pp. 73-88). Edward Elgar Publishing.

Boutros, M., Ben-David, I., Graham, J. R., Harvey, C. R., & Payne, J. W. (2020) "The persistence of miscalibration," *National Bureau of Economic Research Working Paper,* 28010.

Brown, B. W., & Maital, S. (1981) "What do economists know? An empirical study of experts' expectations," *Econometrica*,49(2), 491-504.

Campitelli, G., & Gerrans, P. (2014) "Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach," *Memory & Cognition*, 42(3), 434-447.

Carlé, T. A., Lahav, Y., Neugebauer, T., & Noussair, C. N. (2019). Heterogeneity of beliefs and trade in experimental asset markets. *Journal of Financial and Quantitative Analysis*, 54(1), 215-245.

Carroll, C. D. (2003) "Macroeconomic expectations of households and professional forecasters," *Quarterly Journal of Economics*, 118(1), 269-298.

Chi, M. T. (2006). Two approaches to the study of experts' characteristics. The *Cambridge Handbook of Expertise and Expert Performance*, 21-30.

Chen, Z., Gaba, A., Tsetlin, I., & Winkler, R. L. (2022). Evaluating quantile forecasts in the M5 uncertainty competition. *International Journal of Forecasting*, forthcoming.

Cipriani, M., & Guarino, A. (2009) "Herd behavior in financial markets: an experiment with financial market professionals," *Journal of the European Economic Association*, 7(1), 206-233.

Cipriani, M., and Antonio Guarino (2005) "Herd behavior in a laboratory financial market," *American Economic Review*, 95, 1427-1443.

Clement, M.B. (1999). "Analyst Forecast Accuracy: Do Ability, Resources, and Portfolio Complexity Matter?," *Journal of Accounting and Economics* 27, 285-303.

Clement, M.B. and S.Y. Tse (2005). "Financial Analyst Characteristics and Herding Behavior in Forecasting," *Journal of Finance* 60, 307-341.

Cohn, A. J., Fehr, E., & Maréchal, M. A. (2015) “Evidence for countercyclical risk aversion: An experiment with financial professionals,” *American Economic Review*, 105(2), 860-85.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. Judgment and Decision making, 7(1), 25-47.

Corgnet, B., Desantis, M., & Porter, D. (2018) “What makes a good trader? On the role of intuition and reflection on trader performance,” *Journal of Finance*, 73(3), 1113-1137.

Corgnet, B., Deck, C., DeSantis, M., & Porter, D. (2022) “Forecasting Skills in Experimental Markets: Illusion or Reality?,” *Management Science*, 68(7), 5216-5232.

Cowles, A. (1933) “Can stock market forecasters forecast?” *Econometrica*, 1(3), 309-324.

Da, Z., & Huang, X. (2020) “Harnessing the wisdom of crowds,” *Management Science*, 66(5), 1847-1867.

Daniel, K., Hirshleifer D. & Subrhamanyam, A. (1998) “Investor psychology and security market under‐and overreactions,” *Journal of Finance* 53 (6), 1839-1885.

Deaves, R., Lüders, E., & Schröder, M. (2010) “The dynamics of overconfidence: Evidence from stock market forecasters,” *Journal of Economic Behavior & Organization*, 75(3), 402-412.

De Bondt, W. (1991) “What do economists know about the stock market?” *Journal of Portfolio Management*, 17(2), 84-90.

De Bondt, W. (1993) “Betting on trends: Intuitive forecasts of financial risk and return,” *International Journal of Forecasting*, 9(3), 355-371.

Dokko, Y., & Edelstein, R. H. (1989) “How well do economists forecast stock market prices? A study of the Livingston surveys,” *American Economic Review*, 79(4), 865-871.

Engelberg, J., Sasseville, C., & Williams, J. (2012) “Market madness? The case of mad money,” *Management Science*, 58(2), 351-364.

Easterwood, J. C. and S. R. Nutt (1999) “Inefficiency in analysts’ earnings forecasts: Systematic misreaction or systematic optimism?,” *Journal of Finance* 54, 1777-1797.

Fernandes, D., Lynch Jr, J. G., & Netemeyer, R. G. (2014) “Financial literacy, financial education, and downstream financial behaviors,” *Management Science,* 60(8), 1861-1883.

Fernandez-Perez, A., Garel, A., & Indriawan, I. (2020) “Natural gas storage forecasts: Is the crowd wiser?” *The Energy Journal*, 41(5), 213-238.

Fréchette G., (2015) “Laboratory experiments: Students vs. professionals,” Ch. 17 *in Handbook of Experimental Economic Methodology*, G. Fréchette and A. Schotter eds., Oxford University Press, Oxford, United Kingdom.

Frederick, S. (2005) “Cognitive reflection and decision making,” *Journal of Economic Perspectives*, 19(4), 25-42.

Finucane, M. L., & Gullion, C. M. (2010) “Developing a tool for measuring the decision-making competence of older adults,” *Psychology and Aging*, 25(2), 271.

Genesove, D., & Mayer, C. (2001) "Loss aversion and seller behavior: Evidence from the housing market," *Quarterly Journal of Economics*, 116(4), 1233-1260.

Gaba A., Tsetlin I., Winkler R. L. (2017). "Combining Interval Forecasts," *Decision Analysis*, 14(1), (pp1-20).

Gay, G. D., B. J. Simkins, & M. Turac (2009) "Analyst forecasts and price discovery in futures markets: The case of natural gas storage," *Journal of Futures Markets* 29, 451-477.

Gervais, S. and Odean, T. (2001) "Learning to be overconfident," *Review of Financial Studies* 14(1), 1-27.

Glaser, M., Iliewa, Z., & Weber, M. (2019) "Thinking about prices versus thinking about returns in financial markets," *The Journal of Finance*, 74(6), 2997-3039.

Glaser, M., Langer, T., & Weber, M. (2007) "On the trend recognition and forecasting ability of professional traders," *Decision Analysis*, 4(4), 176-193.

Graham, J. R. (1999) "Herding among Investment Newsletters: Theory and Evidence," *The Journal of Finance*, 54(1), 237-268.

Graham, J. (2022) "Presidential Address: Corporate Finance and Reality," *National Bureau of Economic Research Working Paper*, 29841.

Haigh, M. S., & List, J. A. (2005) "Do professional traders exhibit myopic loss aversion? An experimental analysis," *The Journal of Finance*, 60(1), 523-534.

Hanaki, N. (2022) "Risk misperceptions of structured financial products with worst-of payout characteristics revisited," *Journal of Experimental and Behavioral Economics*, 33, 100604.

Harrison, G. W., & List, J. A. (2004) "Field experiments," *Journal of Economic Literature*, 42(4), 1009-1055.

Hefti, A., Heinke, S., & Schneider, F. (2018). Mental capabilities, heterogeneous trading patterns and performance in an experimental asset market. Working Paper.

Holm, S. (1979). "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics,* 6 (2), 65-70.

Holmen, M., Holzmeister, F., Kirchler, M., Stefan, M., Wengstrom, E. (2023) "Economic Preferences and Personality Traits among Finance Professionals and the General Population," *Economic Journal*, 133 (November), 2949-2977.

Holzmeister, F., Huber, J., Kirchler, M., Lindner, F., Weitzel, U., & Zeisberger, S. (2020) "What drives risk perception? A global survey with financial professionals and laypeople," *Management Science*, 66(9), 3977-4002.

Hong, H. and J. D. Kubik (2003) "Analyzing the analysts: Career concerns and biased earnings forecasts," *Journal of Finance* 58, 313-351.

Jain, K., Mukherjee, K., Bearden, J. N., & Gaba, A. (2013). "Unpacking the future: A nudge toward wider subjective confidence intervals." *Management Science*, 59(9), 1970-1987.

Kahneman, D. (2011) *Thinking, Fast and Slow*, Macmillan.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021) *Noise: A Flaw in Human Judgment*, Little, Brown.

Kaustia, M., Alho, E. & Puttonen, V. (2008) "How much does expertise reduce behavioral biases? The case of anchoring effects in stock return estimates," *Financial*

Management, 37 (3), 391-412

Kirchler, M., Lindner, F., & Weitzel, U. (2018) "Rankings and risk-taking in the finance industry," *The Journal of Finance*, 73(5), 2271-2302

Kirchler, M., Lindner, F., & Weitzel, U. (2020) "Delegated investment decisions and rankings," *Journal of Banking & Finance*, 120, 105952.

Lakonishok, J. (1980) "Stock market return expectations: Some general properties," *Journal of Finance*, 35(4), 921-931.

Laster, D., Bennett, P., & Geoum, I. S. (1999) "Rational bias in macroeconomic forecasts," *Quarterly Journal of Economics*, 114(1), 293-318.

Lee, D. S. (2009) "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *Review of Economic Studies*, 76(3), 1071-1102.

Lim, T. (2002) "Rationality and analysts' forecast bias." *Journal of Finance* 56, 369-385.

Lindner, F., Kirchler, M., Rosenkranz, S., Weitzel, U. (2021) "Social motives and risk-taking in investment decisions," *Journal of Economic Dynamics and Control*, 127: 104116.

List, J. (2007) "Field experiments: A bridge between lab and naturally occurring data," *The BE Journal of Economic Analysis & Policy*, 5(2), 1-47.

List, J. A. (2020) "Non est disputandum de generalizability? a glimpse into the external validity trial," *National Bureau of Economic Research Working Paper,* (No. w27535).

List, J. (2021) "The Voltage Effect in Behavioral Economics" (No. 00733). The Field

Experiments Website.

List, J. A. (2003) “Does market experience eliminate market anomalies?” *Quarterly Journal of Economics,* 118, 41-71.

List, J. A. (2004). “Neoclassical theory versus prospect theory: Evidence from the marketplace,” *Econometrica*, 72(2), 615-625.

List, J. A., Shaikh, A. M., & Xu, Y. (2019) “Multiple hypothesis testing in experimental economics,” *Experimental Economics*, 22(4), 773-793.

Lo, A. W., & Hasanhodzic, J. (2010) *The Heretics of Finance: Conversations with Leading Practitioners of Technical Analysis,* John Wiley & Sons.

Lo, A. W., & Hasanhodzic, J. (2010) *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals*, John Wiley & Sons.

Lo, A. W., Mamaysky, H., & Wang, J. (2000) “Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation,” *Journal of Finance*, *55*(4), 1705-1765.

Locke, P. R., and Mann, S.C. (2005) “Professional trader discipline and trade disposition,” *Journal of Financial Economics,* 76 (2), 401-444.

Menkhoff, L., (2010), The use of technical analysis by fund managers: International evidence, *Journal of Banking & Finance*, 34, 11, 2573-2586

Noussair, C. N., Tucker, S., & Xu, Y. (2016). Futures markets, cognitive ability, and mispricing in experimental asset markets. *Journal of Economic Behavior & Organization*, 130, 166-179.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009) “Cognitive abilities and behavioral biases,” *Journal of Economic Behavior & Organization*, 72, 147-152.

Pearce, D. K. (1984) “An empirical analysis of expected stock price movements,” *Journal of Money, Credit and Banking*, 16(3), 317-327.

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012) “Analytic cognitive style predicts religious and paranormal belief,” *Cognition*, 123, 335-346.

Pierdzioch, C., & Rülke, J. C. (2012) “Forecasting stock prices: Do forecasters herd?” *Economics Letters*, 116(3), 326-329.

Pierdzioch, C., J.C. Rülke, & G. Stadtmann (2013) “Forecasting metal prices: Do forecasters herd?” *Journal of Banking and Finance* 37, 150-158.

Razen, M., Kirchler, M., & Weitzel, U. (2020) “Domain-specific risk-taking among finance professionals,” *Journal of Behavioral and Experimental Finance*, 27, 100331.

Ryan, R. M. (1982) “Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory,” *Journal of Personality and Social Psychology*, 43, 450-461.

Scharfstein, D. S., & Stein, J. C. (1990) “Herd behavior and investment,” *American Economic Review*, 80 (3), 465-479.

Schwaiger, R., Kirchler, M., Lindner, F., & Weitzel, U. (2020) “Determinants of investor expectations and satisfaction. A study with financial professionals,” *Journal of Economic Dynamics and Control*, 110, 103675.

Shenhav, A., Rand, D. G., & Greene, J. D. (2012) “Divine intuition: Cognitive style influences belief in god,” *Journal of Experimental Psychology: General*, 141, 423-428.

Snowberg, E., & Yariv, L. (2021) “Testing the waters: Behavior across participant pools,” *American Economic Review*, 111(2), 687-719.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014) “Tweets and trades: The information content of stock microblogs,” *European Financial Management*, 20(5), 926-957.

Steinmayr, A. (2020). “MHTREG: Stata module for multiple hypothesis testing controlling for FWER,” Working Paper.

Tetlock, P. E. (2009a) *Expert political judgment: How Good Is It? How Can We Know?* Princeton University Press.

Tetlock, P. (2009b) “Reading Tarot on K Street,” *The National Interest*, (103), 57-67.

Tetlock, P. E., & Gardner, D. (2016) *Superforecasting: The art and science of prediction.* Random House.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011) “The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks,” *Memory & Cognition*, 39(7), 1275-1289.

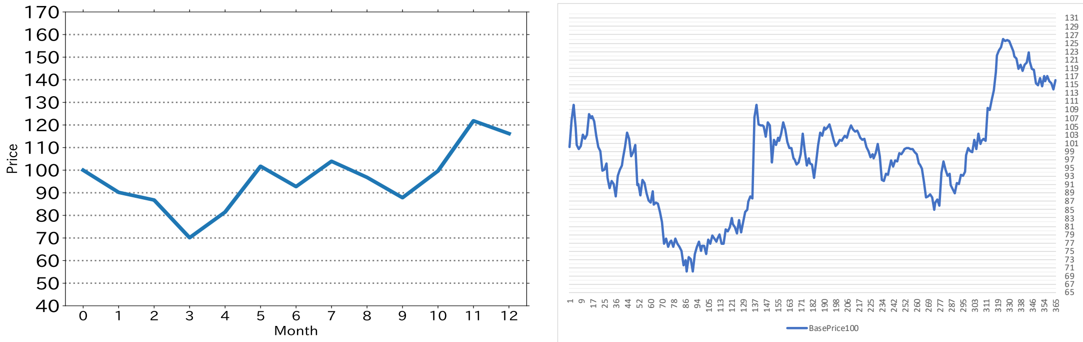
Toplak, M. E., West, R. F., & Stanovich, K. E. (2014) “Assessing miserly information processing: An expansion of the Cognitive Reflection Test,” *Thinking & Reasoning*, 20(2), 147-168.

Weinstein, N. D. (1980). "Unrealistic optimism about future life events." *Journal of Personality and Social Psychology*, 39(5), 806-820.

Weitzel, U., Huber, C., Huber, J., Kirchler, M., Lindner, F., & Rose, J. (2020) "Bubbles and financial professionals," *Review of Financial Studies*, 33(6), 2659 - 2696.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013) "Development and testing of an abbreviated numeracy scale: A Rasch analysis approach," *Journal of Behavioral Decision Making*, 26(2), 198-212.

Figures



(a) SPmonthly (Series 1) (b) SPdaily (Series 1)

Figure 1. Series 1 for SPmonthly price time series in panel (a) and SP daily price time series in panel (b). The vertical axis is the value of the index, while the horizontal axis is the period.

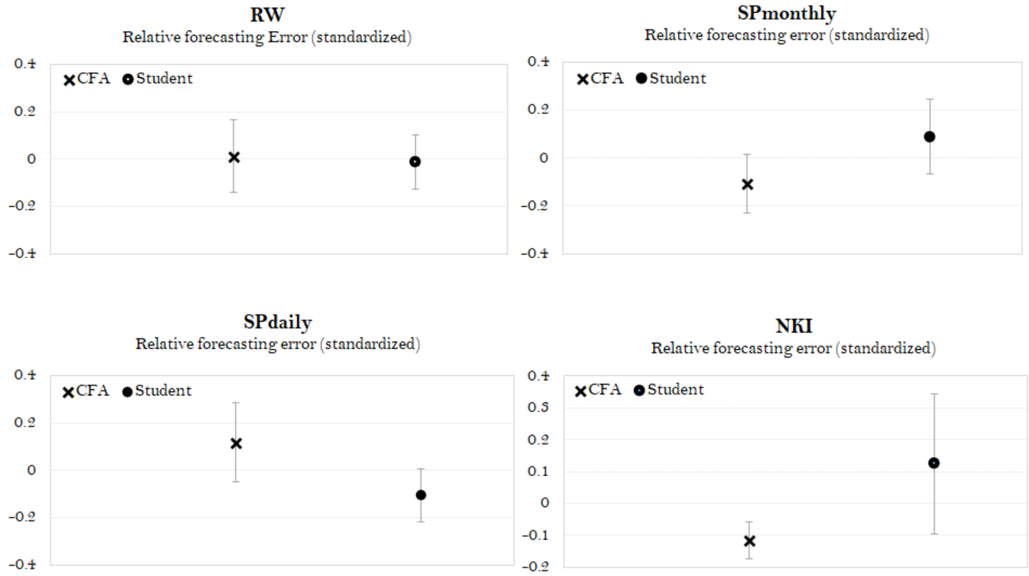


Figure 2. Average standardized RFE (computed at the participant level) for each task across CFAs and students along with 95% confidence intervals.

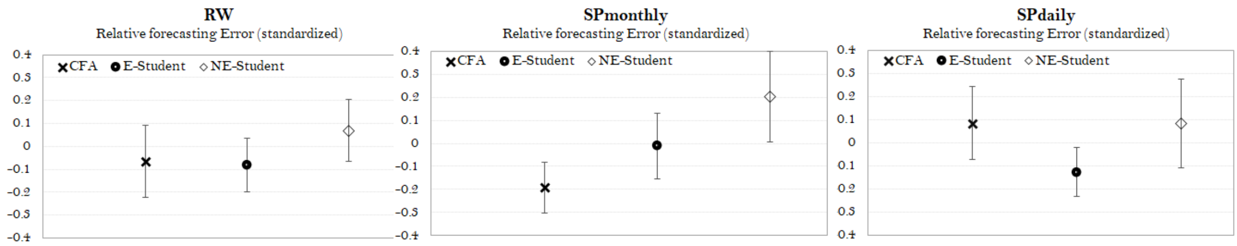


Figure 3. Average standardized RFE (computed at the participant level) for each task for CFAs and the two student samples along with 95% confidence intervals.

Tables

Table 1. Description of forecasting tasks

Tasks	Series	Display	Information	Nature of task	Time frame	CFA Body of Knowledge	Adapted from	Number of Observations	
								CFAs	Students
RW	Artificial	Unnatural	Series	Lab	Static	II.B Probability basics	Bloomfield and Hales (2002)	212	228
SPmonthly (congruent)	Real	Unnatural	Series	Lab	Static	II.C Data visualization	Glaser, Iliewa and Glaser (2019) [Studies 1 & 2]	87	105
SPmonthly (incongruent)	Real	Unnatural	Series (incongruent)					87	101
SPdaily	Real	Natural	Series	Lab	Static	X.G Technical analysis	This paper	198	221
NKI	Real	Natural	Complete	Field	Longitudinal	X.G & III.J Capital Market Expectations	Glaser, Iliewa and Glaser (2019) [Study 3]	717 (Cumulative Total. See Table 2 for the breakdown)	538
Cumulative Total Number of Observations								1,301	1,193

RW means predicting random walk. SP monthly means predicting SP monthly data. SP daily means predicting SP daily data. NKI means predicting the Nikki index.

Table 2. Description of Nikkei forecasting task (NKI)

Wave	Forecast window	Forecasting day	CFAs	Students
1	March 23-26, 2021	April 23, 2021	125	57*
2	Apr. 27-30, 2021	May 21, 2021	113	113
3	May 25-28, 2021	June 25, 2021	125	109
4	June 28-July 2, 2021	July 30, 2021	124	96
5	August 3-6, 2021	August 27, 2021	111	84
6	August 31-September 4, 2021	September 24, 2021	119	79
Cumulative Total Number of Observations			717	538

* The number of participating students was lower than for previous tasks in March because this corresponds to the graduation period in Japan.

Table 3. The difference between relative forecast errors of CFAs and students in four experiments

Dependent variable	Standardized RFE				
Task	RW (1)	RW (1')	SPmonthly (2)	SPdaily (3)	NKI♠ (4)
CFA Dummy	0.005	0.011	-0.039**	0.055**	-0.232**
	(0.075)	(0.037)	(0.017)	(0.027)	(0.100)
CRT score (std)	-0.099**	-0.094**	-0.011	-0.059	0.0006
	(0.045)	(0.045)	(0.009)	(0.034)	(0.028)
Risk aversion (number of safe options, Holt and Laury, 2002)		0.043			
		(0.035)			
Constant	0.251***	0.253***	-0.013	0.046	-0.103
	(0.064)	(0.057)	(0.059)	(0.054)	(0.172)
R^2	0.030	0.032	0.164	0.445	0.053
Observations	7,040	7,040	7,400	8,138	1,181

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses. Series and order fixed effects included, and month for NKI.
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05
CFA Dummy takes value one for participants who are CFAs.
♠ In NKI, we face selection issues as the study takes place over six months, and not all students and CFAs make a forecast each month. We thus estimated Lee bounds (Lee, 2009) in Table C3 in Internet Appendix C (see regression (1)) to correct for selection issues. Both the lower and upper bounds of the CFA Dummy coefficient are negative, and this might not be surprising given that the level of participation is relatively high on average (64.2%).
(std) stands for standardized deviation.

Table 4. The difference between relative forecast errors of CFAs and students in incongruent and congruent treatments of SPmonthly

Dependent variable	Standardized RFE	
Task	SPmonthly (Incongruent) (1)	SPmonthly (Congruent) (2)
CFA Dummy	-0.062**	-0.002
	(0.025)	(0.023)
CRT score (std)	-0.006	-0.024
	(0.011)	(0.015)
Constant	-0.092	0.008
	(0.077)	(0.082)
R^2	0.202	0.147
Observations	3,660	3,740

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses. Series, and order fixed effects included.
Robust standard errors in parentheses. *** *p*-value < 0.01, ** 0.01< *p*-value < 0.05
(std) stands for standardized.

Table 5. Descriptive statistics and comparisons across samples for all variables

				RFE			
Mean <Median> (Standard Deviation)	CRT	Financial literacy	Risk aversion[+]	RW	SPmonthly	SPdaily	NKI
CFA Japan (n = 212, CRT) & (n = 110, Financial literacy)	5.127	10.927	4.142	40.843	7,708	8.191	5.002
	<6>	<11>	<4>	<20>	<22.122>	<5.715>	<4.658>
	(1.227)	(1.254)	(1.097)	(40.657)	(85,487)	(8.522)	(2.500)
E-Student (n = 228)	5.316	8.285	4.263	40.122	12,329	7.689	6.632
	<6>	<9>	<4>	<40>	<22.840>	<5.560>	<4.582>
	(0.899)	(2.448)	(1.154)	(35.504)	(125,112)	(7.968)	(9.272)
NE-Student (n = 149)	4.181	7.040	4.409	48.813	19,559	8.187	
	<5>	<7>	<4>	<40>	<33.863>	<5.557>	
	(1.586)	(2.379)	(1.461)	(35.937)	(151,302)	(8.791)	
Wilcoxon Rank Sum Test (p-value)							
CFA vs E-Student	0.396	<0.001	0.160	0.550	0.013	0.034	0.184
CFA vs NE-Student	<0.001	<0.001	0.005	<0.001	<0.001	0.300	
CFA vs All students	0.009	<0.001	0.021	<0.001	<0.001	0.053	
E-Student vs NE-Student	<0.001	<0.001	0.099	<0.001	<0.001	0.462	

Note**:** + Number of safe options (Holt and Laury, 2002)
E-student means students from the elite university. While NE-student means students from the non-elite university.

Table 6. Relative Forecast Error and CRT as a mediator in RW, SPmonthly, SPdaily

Dependent variable	Standardized RFE					
Task	RW		SPmonthly		SPdaily	
	(1)	(2)	(3)	(4)	(5)	(6)
CFA Dummy	0.019	0.013	-0.037**	-0.039**	0.064**	0.057**
	(0.075)	(0.075)	(0.017)	(0.017)	(0.029)	(0.027)
NE Student Dummy	0.230***	0.193**	0.052**	0.038	0.048	0.012
	(0.068)	(0.075)	(0.022)	(0.023)	(0.032)	(0.036)
CRT score (std)	-	-0.043	-	-0.016**	-	-0.042
		(0.031)		(0.008)		(0.022)
Constant	0.194***	0.206***	0.020	0.022	0.006	0.017
	(0.060)	(0.060)	(0.055)	(0.055)	(0.047)	(0.046)
CFA vs NE Student (*p*-value)	0.007	0.030	<0.001	<0.001	0.664	0.297
R^2	0.032	0.034	0.196	0.196	0.434	0.435
Observations	9,424	9,424	10,080	10,080	10,798	10,798

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses. Series and order fixed effects included.
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05
NE-Student Dummy takes value one for participants who belong to the non-elite university.
(std) stands for standardized.

Table 7. Relative Forecast Error and financial literacy as a mediator in RW, SPmonthly, and SPdaily

Dependent variable	Standardized RFE											
Task	RW				SPmonthly				SPdaily			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CFA Dummy	0.051	0.046	0.127	0.122	-0.043**	-0.045**	-0.035	-0.043**	0.032	0.030	0.042	0.038
	(0.095)	(0.094)	(0.100)	(0.100)	(0.018)	(0.018)	(0.019)	(0.020)	(0.026)	(0.026)	(0.030)	(0.030)
NE Student Dummy	0.230***	0.205***	0.194***	0.186**	0.052**	0.036	0.049**	0.035	0.048	0.039	0.043	0.038
	(0.068)	(0.076)	(0.072)	(0.078)	(0.022)	(0.023)	(0.023)	(0.024)	(0.032)	(0.032)	(0.034)	(0.034)
CRT score (std)	-	-0.029	-	-0.011	-	-0.020**	-	-0.019**	-	-0.010	-	-0.008
		(0.033)		(0.033)		(0.009)		(0.009)		(0.011)		(0.012)
Financial literacy (std) [a]	-	-	-0.076**	-0.073	-	-	-0.007	-0.002	-	-	-0.010	-0.007
			(0.036)	(0.037)			(0.010)	(0.010)			(0.017)	(0.018)
Constant	0.177***	0.185***	0.171***	0.174***	-0.016	-0.014	-0.017	-0.014	-0.030	-0.027	-0.031	-0.029
	(0.063)	(0.063)	(0.063)	(0.063)	(0.054)	(0.054)	(0.054)	(0.054)	(0.043)	(0.043)	(0.043)	(0.043)
CFA vs NE Student (*p*-value)	0.066	0.118	0.557	0.576	<0.001	<0.001	0.002	0.003	0.627	0.791	0.968	0.988
R^2	0.033	0.034	0.037	0.037	0.198	0.198	0.198	0.198	0.459	0.459	0.459	0.459
Observations	7,792	7,792	7,792	7,792	8,840	8,840	8,840	8,840	9,238	9,238	9,238	9,238

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses. Series and order fixed effects included.
Robust standard errors in parentheses. *** *p*-value < 0.01, ** 0.01< *p*-value < 0.05
NE-Student Dummy takes value one for participants who belong to the non-elite university.
(std) stands for standardized.
[a] Collinearity does not seem to be an issue as the variance inflation factors associated with the three regressors (CFA Dummy, CRT and Financial literacy) are below 1.45, regardless of the regression.

INTERNET APPENDIX

Internet Appendix A. Instructions

A.1. RW

A.1.1. Instructions

We have constructed a model of a random process that works much like flipping a fair coin. Using this model, we have created sequences of outcomes. An upward movement indicates a "heads" outcome, and a downward movement indicates a "tails" outcome.

Since outcomes of coin flips are unpredictable, they result in a sequence known as a "random walk." That is, statistical models are unable to predict future outcomes from past ones and, on average, there is no upward or downward trend. Random walk sequences almost always contain intervals of recognizable patterns. However, since these patterns can change greatly at any time, statistical models are still unable to predict future outcomes.

You will be shown 16 plots we have created as described. You are given one unit of bet that will generate either 0 or 100 points depending on the next movement (on how the series move next). Your bet generates 100 points if it moves upward, and 0 point if it moves downward.

You are asked to name the price (in points) at which you are willing to sell this bet back to us. The price you can set is between 0 and 100 points.

One in ten participants will be selected for financial reward. If you are selected for the financial reward, we will calculate your earnings for this experiment as follows.

We will select one of the 16 plots at random and then generate a random number between 0 and 100 (each number is equally likely). If the randomly generated number is less than or equal to the price you have set for the bet, then, you will keep the bet and your payoff will be either 0 or 100 points depending on the next move of the series represented on the plot. If the randomly generated number is greater than the price you have set for the bet, then, you will receive the points equal to the randomly generated number.

<u>**Example 1**</u> If you set the price for the bet to be 50 points and the randomly generated number is 40, you keep the bet, and your payoff will be either 0 or 100 points.

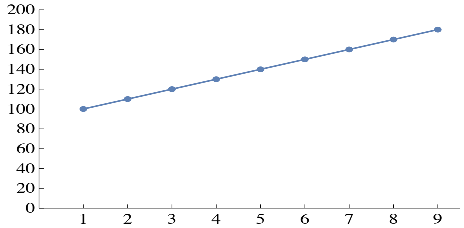
<u>**Example 2**</u> If you set the price for the bet to be 50 points and the randomly generated number is 60, you receive 60 points.

You will be paid for the points you have obtained in the selected plot.1 point will be converted into 100/20 JPY. The payment will be in the form of an Amazon gift card. You will not be informed about the accuracy of your forecast until the end of the experiment.

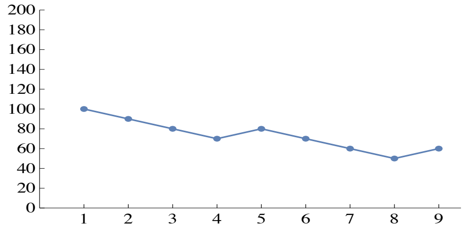
A.1.2. Figures

Series used in RW. As is in Bloomfield and Hales (2002), Series 9 to 16 are the mirror images of Series 1 to 8, respectively.

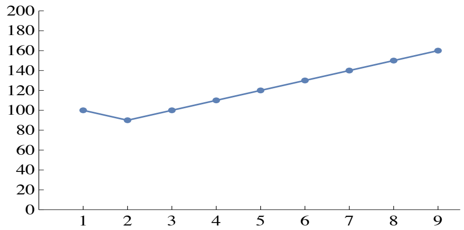
Series 1



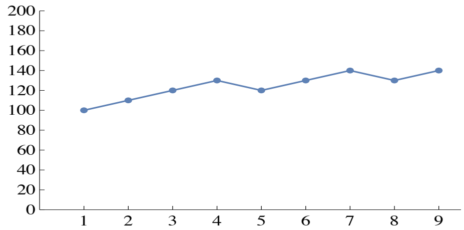
Series 5



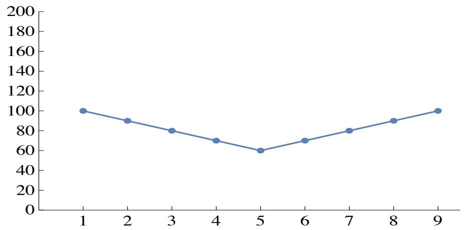
Series 2



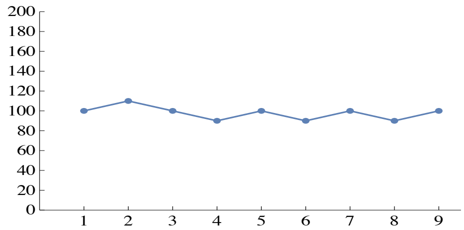
Series 6



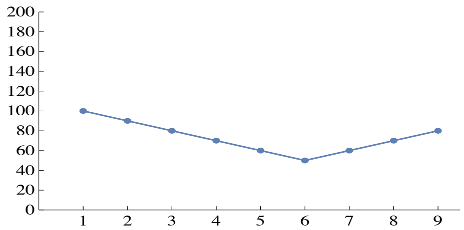
Series 3



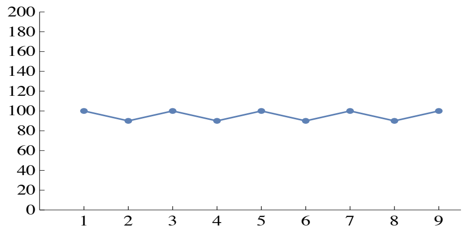
Series 7



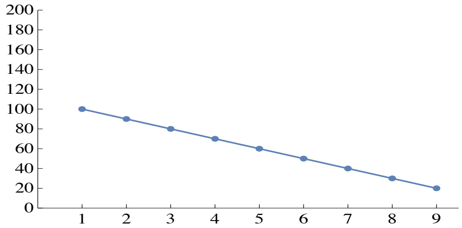
Series 4



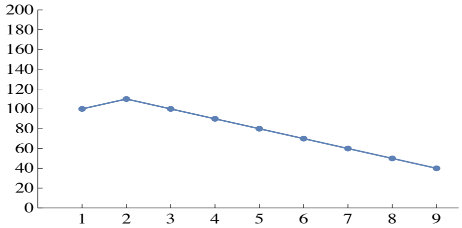
Series 8



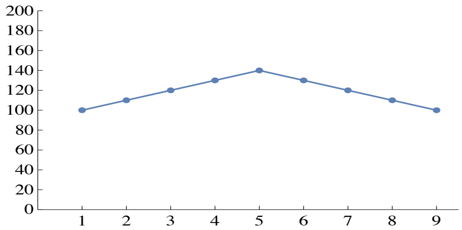
Series 9



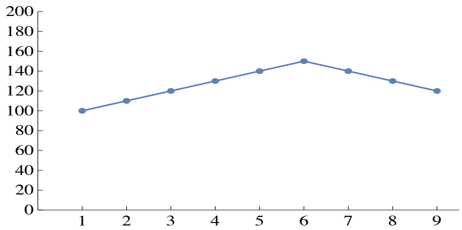
Series 10



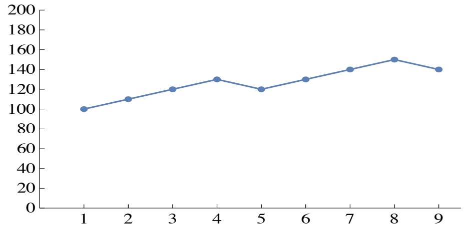
Series 11



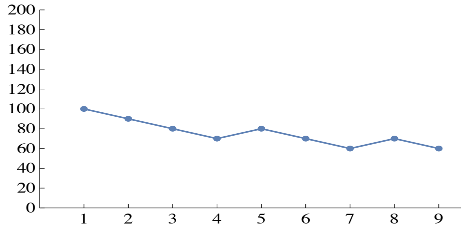
Series 12



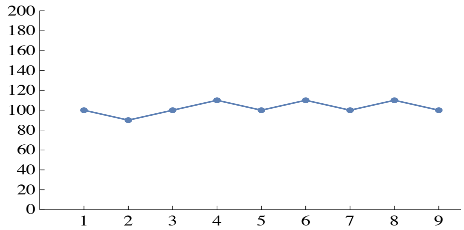
Series 13



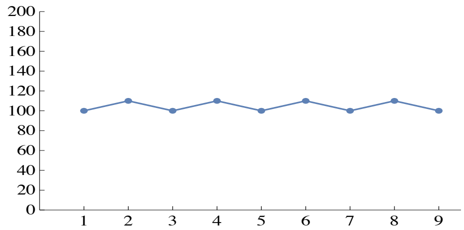
Series 14



Series 15



Series 16



A.2. SPmonthly

A.2.1. Instructions

Congruent treatment with price predictions

In this experiment, there are 20 tasks. In each task, you will be shown 20 graphs showing 12 months of end-of-month prices of a stock, and **asked to forecast what will be the end-of-month price for this stock 1 month after the last price shown on the graph**.

Each graph shows the price movement of a randomly selected stock from the S&P 500 starting from a randomly selected month between January 2008, and June 2018. You will not be told about the name of the stock or the starting date, which was randomly selected.

Please note that end-of-day prices have been rescaled so that all starting prices will be equal to 100.

You will be rewarded based on the accuracy of your forecasts as follows.

$$max\,\{200 - 1000 \times \left|\frac{price\ forecast - realized\ price}{realized\ price}\right|, 0\}$$

If your forecast is exactly at the realized price, then you will receive 200 points. For each percentage point difference between your forecast and the realized price, 10 points will be subtracted.

If your forecast differs from the realized price by more than 20%, you will receive 0 point.

You will do a similar forecasting task for all the 20 randomly chosen stocks. One in ten participants will be selected for financial reward. If you are selected for the financial reward, one of the 20 predictions will be selected randomly, and you will receive the reward according to the points you have earned in the selected task. 1 point will be converted into 100 / 20 JPY. The payment will be in the form of an Amazon gift card.

You will not be informed about the accuracy of your forecast until the end of the experiment.

Incongruent treatment with returns predictions

In this experiment, there are 20 tasks. In each task, you will be shown 20 graphs showing 12 months of end-of-month prices of a stock and **asked to forecast what will be the return for this stock one month after the last price shown on the graph**. Here, the return is defined as

$$return = \frac{realized\ price - last\ price}{last\ price} \times 100$$

(Only the first month, instead of the closing price of the last month, we use the closing price of the first day of the month).

Each graph shows the price movement of a randomly selected stock from the S&P 500 starting from a randomly selected month between January 2008, and June 2018. You will not be told about the name of the stock or the starting date which was randomly selected. **Please note that end-of-day prices have been rescaled so that all starting prices will be equal to 100.**

You will be rewarded based on the accuracy of your forecasts as follows.

$$max\,\{200 - 10 \times |return\ forecast - realized\ return\ |, 0\}$$

If your forecast is exactly at the realized return, you will receive 200 points. For each percentage point difference between your forecast and the realized return, 10 points will be subtracted.

If your forecast differs from the realized return by more than 20%, you will receive 0 point.

You will do a similar forecasting task for all the 20 randomly chosen stocks. One in ten participants will be selected for financial reward. If you are selected for the financial reward, one of the 20 predictions will be selected randomly, and you will receive the reward according to the points you have earned in the selected task. 1 point will be converted into 100/20 JPY. The payment will be in the form of an Amazon gift card.

You will not be informed about the accuracy of your forecast until the end of the experiment.

Incongruent treatment with price predictions

In this experiment, there are 20 tasks. In each task, you will be shown 20 graphs showing 12 months of monthly return of a stock, and **asked to forecast what will be the price for this stock 30 days after the last return shown on the graph**. Here, the monthly return is defined as

$$return = \frac{realized\ price - last\ price}{last\ price} \times 100$$

(Only the first month, instead of the closing price of the last month, we use the closing price of the first day of the month).

Each graph shows the return movement of a randomly selected stock from the S&P 500 starting from a randomly selected month between January, 2008, and June, 2018. You will not be told about the name of the stock or the starting date, which was randomly selected.

You will be rewarded based on the accuracy of your forecasts as follows.

$$max\ \{200 - 1000 \times \left|\frac{price\ forecast - realized\ price}{realized\ price}\right|, 0\}$$

If your forecast is exactly at the realized price, then you will receive 200 points. For each percentage point difference between your forecast and the realized price, 10 points will be subtracted.

If your forecast differs from the realized price by more than 20%, you will receive 0 point.

To ease your forecasting task, the closing price of the final month is also shown. **Please note that end-of-day prices have been rescaled so that all starting prices will be equal to 100.**

You will do a similar forecasting task for all the 20 randomly chosen stocks. One in ten participants will be selected for financial reward. If you are selected for the financial reward, one of the 20 predictions will be selected randomly, and you will receive the reward according to the points you have earned in the selected task. 1 point will be converted into 100 /20 JPY. The payment will be in the form of an Amazon gift card.

You will not be informed about the accuracy of your forecast until the end of the experiment.

Congruent treatment with returns predictions

In this experiment, there are 20 tasks. In each task, you will be shown 20 graphs showing 12 months of monthly return of a stock, and **asked to forecast what will be the return for this stock 1 month after the last return shown on the graph**. Here, the monthly return is defined as

$$return = \frac{realized\ price - last\ price}{last\ price} \times 100$$

(Only the first month, instead of the closing price of the last month, we use the closing price of the first day of the month）.

Each graph shows the return movement of a randomly selected stock from the S&P 500 starting from a randomly selected month between January 2008, and June 2018. You will not be told about the name of the stock or the starting date, which was randomly selected.

You will be rewarded based on the accuracy of your forecasts as follows.

$$max\ \{200 - 10 \times |return\ forecast - realized\ return\ |, 0\}$$

If your forecast is exactly at the realized return, then you will receive 200 points. For each percentage point difference between your forecast and the realized return, 10 points will be subtracted.

If your forecast differs from the realized return by more than 20%, you will receive 0 point.

To ease your forecasting task, the closing price of the final month is also shown. **Please note that end-of-day prices have been rescaled so that all starting prices will be equal to 100.**

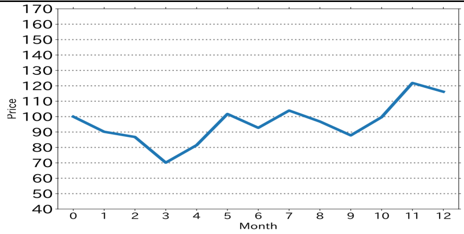
You will do a similar forecasting task for all the 20 randomly chosen stocks. One in ten participants will be selected for financial reward. If you are selected for the financial reward, one of the 20 predictions will be selected randomly, and you will receive the reward according to the points you have earned in the selected task. 1 point will be converted into 100/20 JPY. The payment will be in the form of an Amazon gift card.

You will not be informed about the accuracy of your forecast until the end of the experiment.

A.2.2. Figures

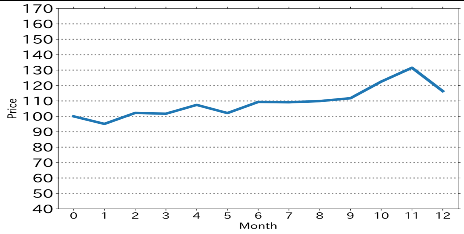
<u>Price series</u>

Series 1

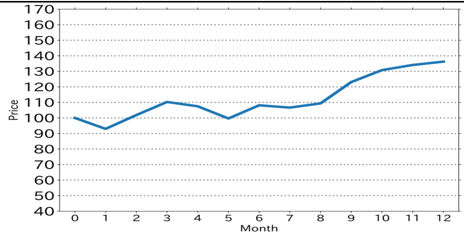


Series 2

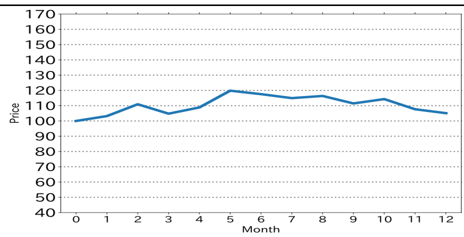
Series 3



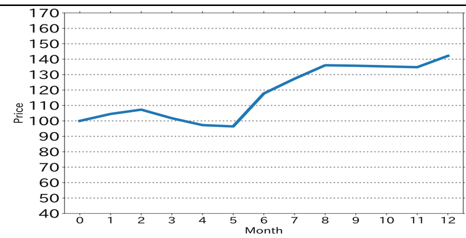
Series 4



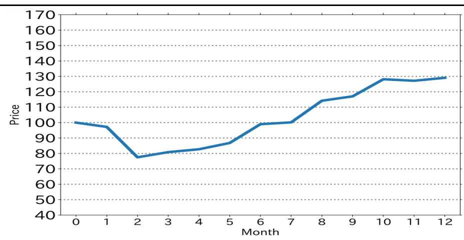
Series 5



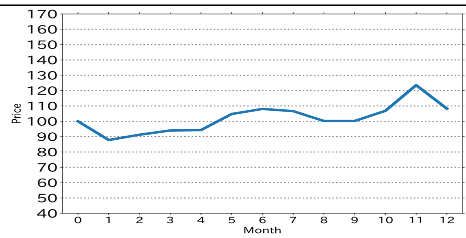
Series 6



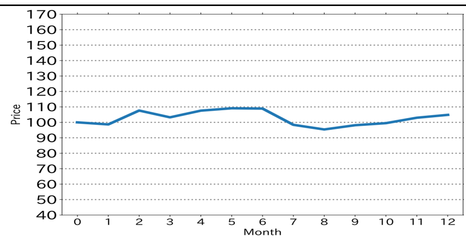
Series 7



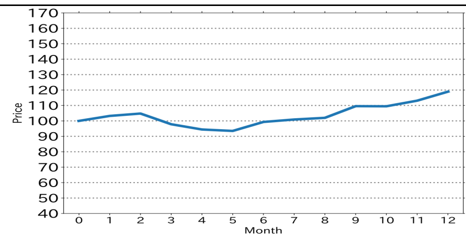
Series 8



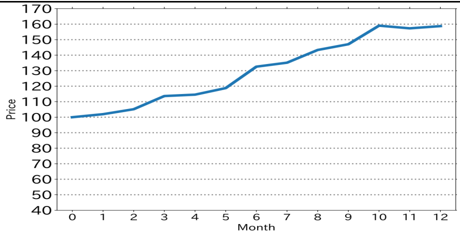
Series 9



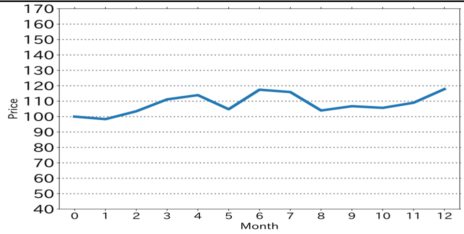
Series 10



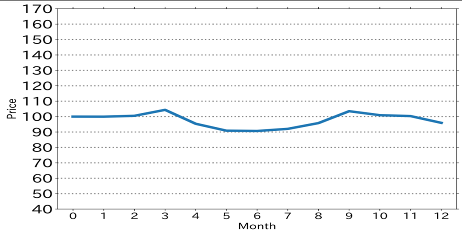
Series 11



Series 12

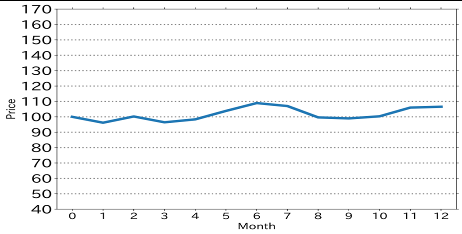


Series 13



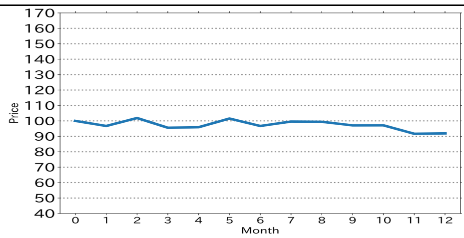
Series 14

Series 15

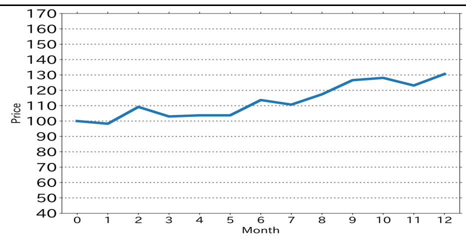


Series 16

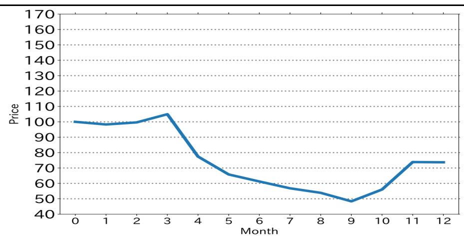
Series 17



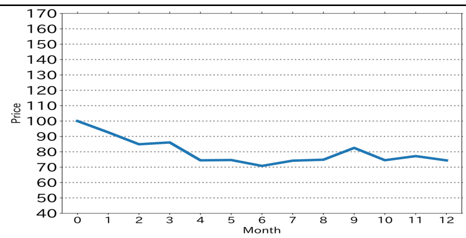
Series 18



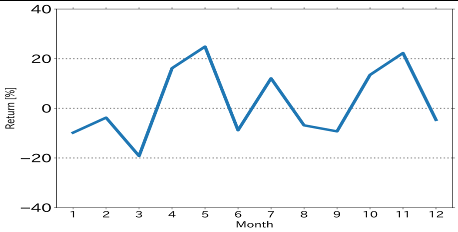
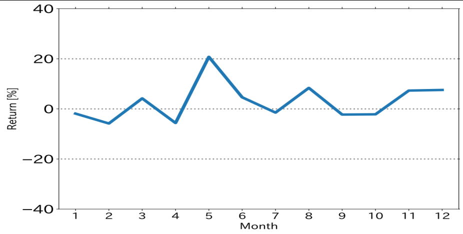
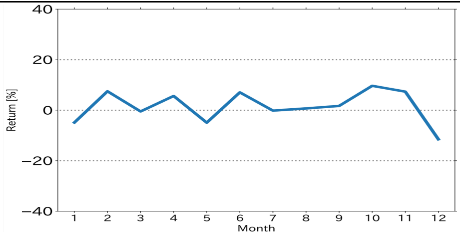
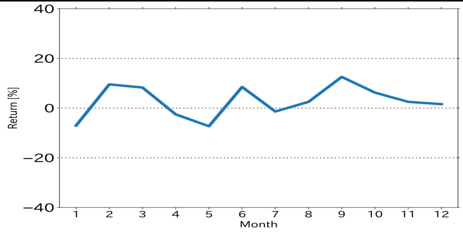
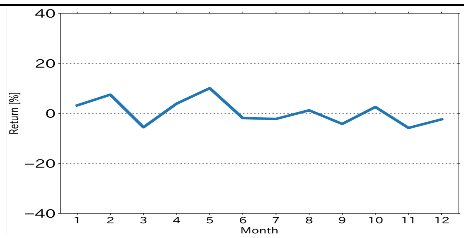
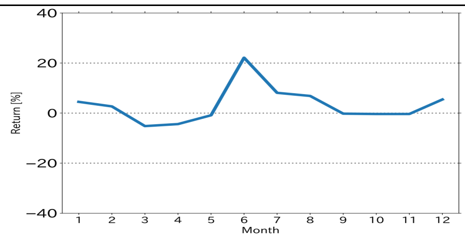
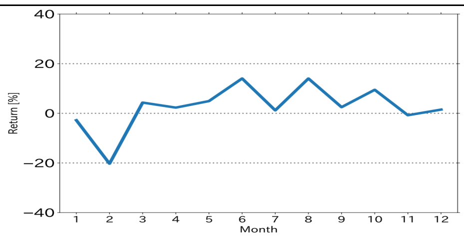
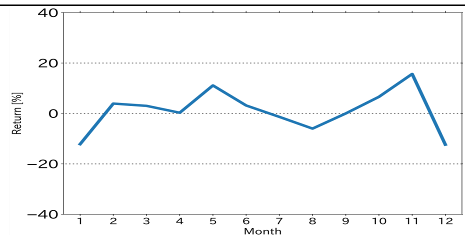
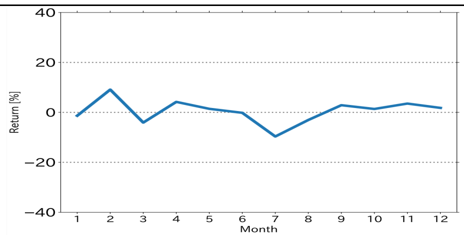
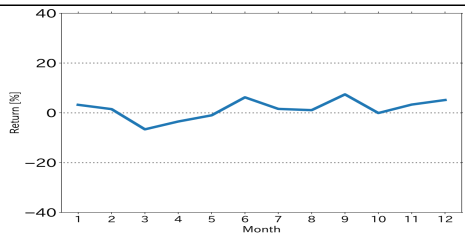
Series 19



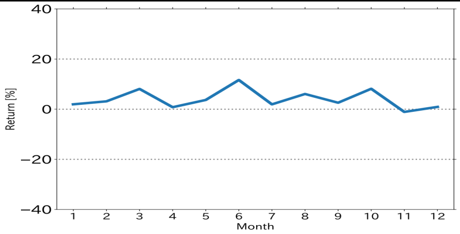
Series 20



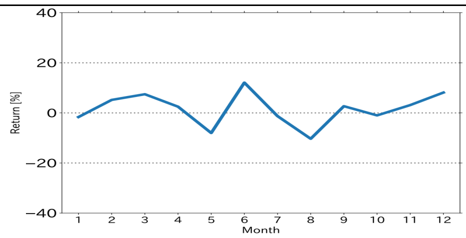
<u>Returns series</u>

Series 1	Series 2
	
Series 3	**Series 4**
	
Series 5	**Series 6**
	
Series 7	**Series 8**
	
Series 9	**Series 10**
	

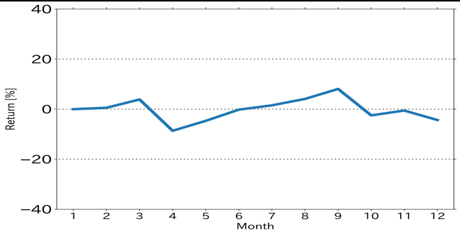
Series 11



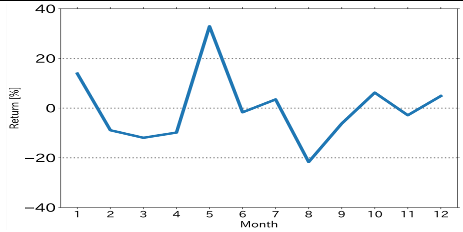
Series 12



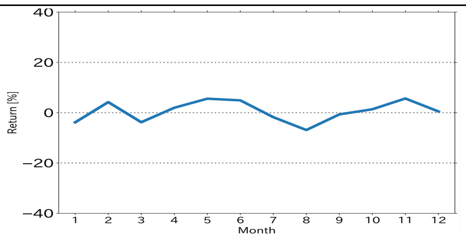
Series 13



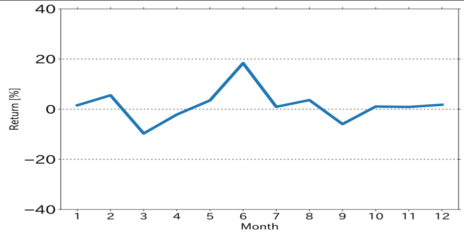
Series 14



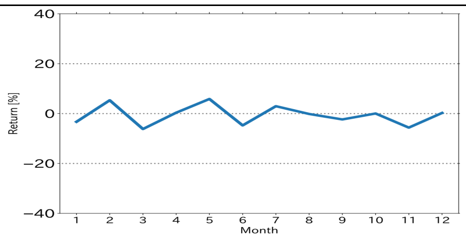
Series 15



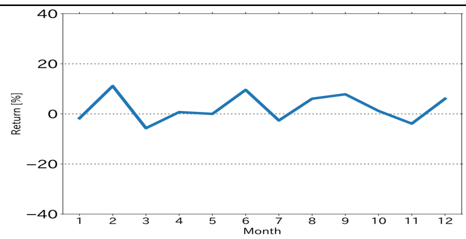
Series 16



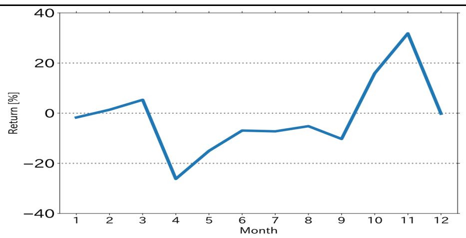
Series 17



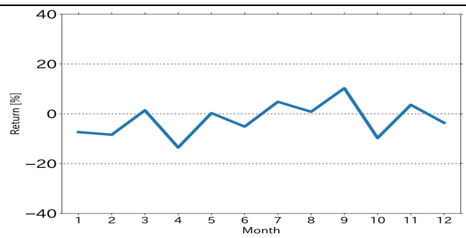
Series 18



Series 19



Series 20



A.3. SPdaily

A.3.1. Instructions

In this experiment, you will be shown 20 graphs showing 12 months of end-of-day prices of randomly selected stocks from the S&P 500 starting from a randomly selected day between January 1st, 2008, and June 30th, 2018. You will not be told about the name of the stock or the starting date, which was randomly selected. **Please note that end-of-day prices have been rescaled so that all starting prices will be equal to 100.**

For each graph, you will be asked to forecast what will be the end-of-day price for this stock **30 days after the last price shown on the graph**.

You will be rewarded based on the accuracy of your forecasts as follows.

$$max\,\{200 - 1000 \times \left|\frac{price\ forecast - realized\ price}{realized\ price}\right|, 0\}$$

If your forecast is exactly at the realized price, then you will receive 200 points. For each percentage point difference between your forecast and the realized price, 10 points will be subtracted.

If your forecast differs from the realized price by more than 20%, you will receive 0 point.

You will do a similar forecasting task for all the 20 randomly chosen stocks. One in ten participants will be selected for financial reward. If you are selected for the financial reward, one of the 20 predictions will be selected randomly, and you will receive the reward according to the points you have earned in the selected task. 1 point will be converted into 100 /20 JPY. The payment will be in the form of an Amazon gift card.

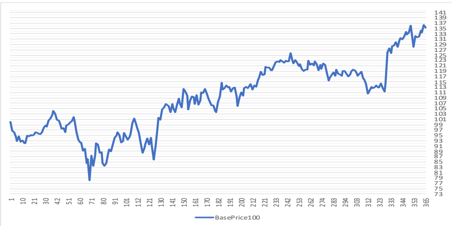
You will not be informed about the accuracy of your forecast until the end of the experiment

A.3.2. Figures

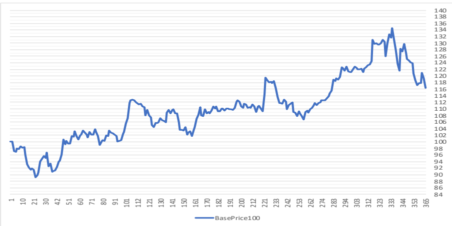
Series used in SPdaily.

Series 1

Series 2

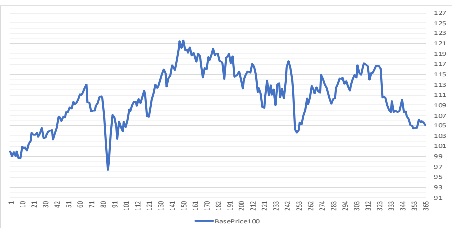


Series 3

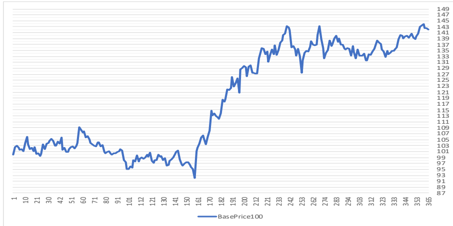


Series 4

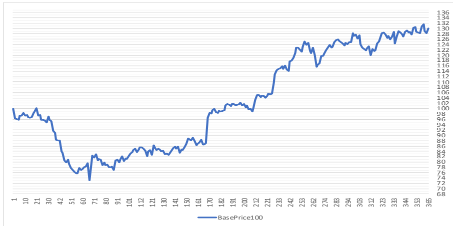
Series 5



Series 6



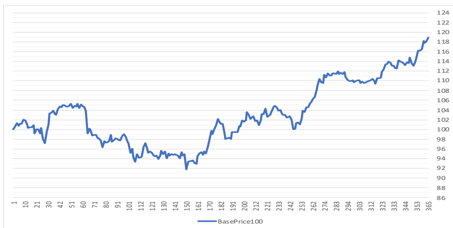
Series 7



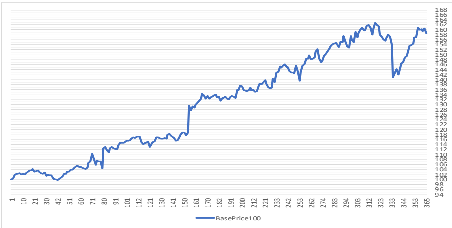
Series 8

Series 9

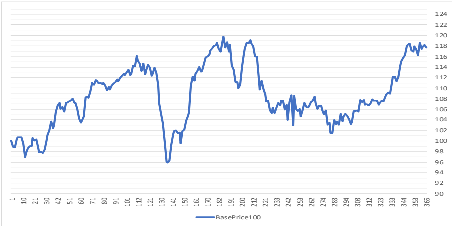
Series 10



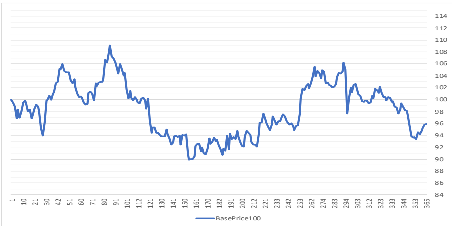
Series 11



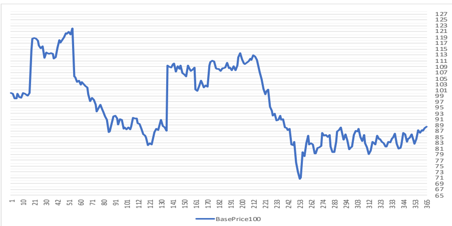
Series 12



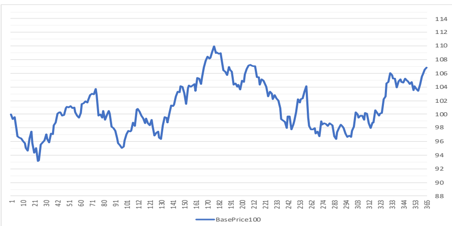
Series 13



Series 14

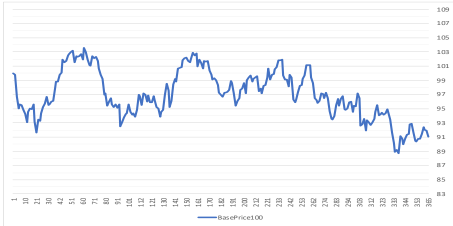


Series 15



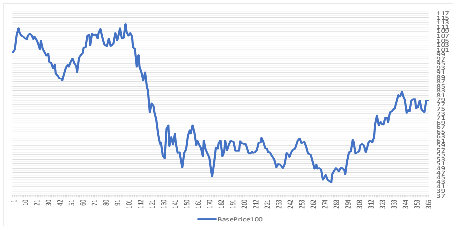
Series 16

Series 17



Series 18

Series 19



Series 20

A.4. NKI.

A.4.1. Forecasting task instructions

Example of feedback and instruction screens.

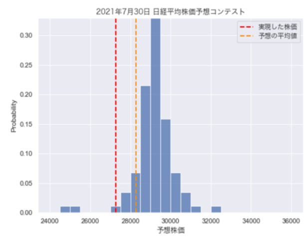
Result of the previous experiment.

Last time, during the period of June 28 to July 2, 2021, we asked you to forecast the closing Nikkei average of Friday, July 30, 2021.

The result is summarized below.

Realized Nikkei average:	**27283.59**
your forecast*:	
The average forecasts:	**28276.84**
The best forecast:	**27410**
The second best forecast:	**27552**
The third best forecast:	**27940**

* It is empty if you did not participate in the last experiment.



Next

Stock Price Forecasting Contest

In this experiment, please forecast the closing Nikkei average of Friday, August 27, 2021.

The participant whose forecast was closest to the realized Nikkei average will be offered a reward of 1000 JPY in the form of Amazon Gift Card (e-mail version). (In case of tie, one will be chosen randomly among the best forecasters.)

Please enter your forecast of the closing Nikkei average of Friday, August 27, 2021, using the slider bar.

The deadline for responding is 23:59 on Friday, August 6, 2021 (Japanese time).

0 4000 8000 12000 16000 20000 24000 28000 32000 36000 40000

Your forecast

Submit

A.4.2. Questionnaire (implemented in Wave 3)

For each of the following statements, please indicate how true it is for you, using the following scale (1-Not at all true, 4-Somewhat true, 7 -Very true):
I put a lot of effort in the forecasting task
I didn't try very hard to do well in the forecasting task
It was important for me to do well at the forecasting task
I didn't put much energy into the forecasting task
I feel this is an important task

How often have you used paid services to help you forecast the Nikkei average?
1-Never, 2-Rarely, 3-Sometimes, 4-Often, 5-Always

In the previous year (2020), how often have you made forecasts about the Nikkei average? 1-Never, 2-Once a year, 3-Once a month, 4-Several times a month, 5-Every day

What is the main strategy you use to forecast the Nikkei average?
1- Use of charts, 2- Use of fundamental information, 3- Trend following, 4-Financial and international news, 5- Intuition, 6- Others (specify), 7-None.

Internet Appendix B. Survey

B.1. Loss attitudes

The loss aversion task is taken from Kirchler et al., (2018). The degree of loss aversion is measured by the number of lotteries chosen out of 6 (the smaller it is, the more loss averse a participant is).

B.1.1 Instructions

In the 6 decisions below, you have to decide whether you want to participate in a lottery where you can win or lose money. For this task, you receive an initial endowment of 18 points.If you reject the lottery, you will only receive your initial endowment.

The initial endowment and one of your 6 decisions below will be randomly selected to calculate your payments. To determine your payment in case you chose the lottery, the program will randomly determine if you receive the loss or the gain. Note that gains and losses are equally likely. Since you do not know which decisions will be selected for payment, and each decisionstands an equal chance of being selected, you should pay attention to the choice you make ineach decision.

Please decide for each of the six rows below.

1. Loss of 3 points with 50% probability or gain of 15 points with 50% probability
2. Loss of 6 points with 50% probability or gain of 15 points with 50% probability
3. Loss of 9 points with 50% probability or gain of 15 points with 50% probability
4. Loss of 12 points with 50% probability or gain of 15 points with 50% probability
5. Loss of 15 points with 50% probability or gain of 15 points with 50% probability
6. Loss of 15 points with 50% probability or gain of 15 points with 50% probability

B.2. Risk attitudes

This task is a variant of the multiple price list of Holt and Laury (2002). The difference is, inour experiments, we fix the probability of the risky outcomes, and the other option is the certain sure amount of which is varied. A similar method is used in He and Hong (2018). The degree of risk aversion is measured by the number of safe options chosen. The higher the number,the more risk-averse a participant is.

B.2.1 Instructions

In the 7 decisions below, you have to decide between two options, A or B. One of your 7 decisions below will be randomly selected, and you will be paid out according to the choice you have made for that selected decision, i.e., either the lottery or the sure payoff. To determine your payment in case you chose the lottery, the program will randomly determine if you receive the lower or the larger amount. Note that the lower and the larger amounts are always equally likely. Since you do not know which decisions will be selected for payment, and each decision stands an equal chance of being selected, you should pay attention to the choice you make in each decision.

1. A: 100% sure amount of 3 points v.s., B: 50% 0 point and 50% 24 points
2. A: 100% sure amount of 6 points v.s., B: 50% 0 point and 50% 24 points
3. A: 100% sure amount of 9 points v.s., B: 50% 0 point and 50% 24 points
4. A: 100% sure amount of 12 points v.s., B: 50% 0 point and 50% 24 points
5. A: 100% sure amount of 15 points v.s., B: 50% 0 point and 50% 24 points
6. A: 100% sure amount of 18 points v.s., B: 50% 0 point and 50% 24 points
7. A: 100% sure amount of 21 points v.s., B: 50% 0 point and 50% 24 points

B.3. Cognitive Reflection Test

We used the 6-question version of the Cognitive Reflection Test (CRT, Frederick, 2005). In particular, we take questions from Finucane and Gullion (2010) and Toplak et al., (2014).

1. If it takes 2 nurses 2 minutes to measure the blood pressure of 2 patients, how long would it take 200 nurses to measure the blood pressure of 200 patients? (in minutes) (Correct answer: 2 minutes; intuitive answer: 200 minutes).

2. A soup and a salad cost 5.50 euros in total. The soup costs 5 euros more than the salad. How much does the salad cost? (in euros). (Correct answer: 0.25 euro; intuitive answer:
0.5 euro)

3. Sally is making sun tea. Every hour, the concentration of the tea doubles. If it takes 6 hours for the tea to be ready, how long would it take for the tea to reach half of the final concentration? (in hours) (correct answer: 5 hours; intuitive answer: 3 hours)

4. If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12days, how long would it take them to drink one barrel of water together? (in days) (correct answer: 4 days; intuitive answer: 9)

5. A man buys a pig for 60 euros, sells it for 70 euros, buys it back for 80 euros, and sells it finally for 90 euros. How much has he made? (correct answer: 20 euros; intuitive answer: 10 euros)

6. Simon decided to invest 8,000 euros in the stock market one day early in 2008. Six months after he invested, on July 17th, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17th to October 17th, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money. (correct answer: c, because the value at this point is 7,000 euros; intuitive response b.)

B.4. Demographics for CFAs

Years of experience:
1: 0-4 years, 2: 5-9 years, 3: 10-14 years, 4: 15-19 years, 5: 20-24 years, 6: 25-29 years, 7: 30-34 years, 8: 35-39 years

Academic degree:
1: Bachelor, 2: Master, 3: Doctor, 4: Others

Age:
1: 25-29, 2:30-34, 3:35-39, 4: 40-44, 5: 45-59, 6: 50-54, 7: 55-59, 8: 60 -64

Sex
1: Female, 2: Male

Annual gross income (million ¥)
1: less than 1, 2: 1-2, 3: 2-4, 4: 4-6, 5: 6-8, 6: 8-10, 7:10-12, 8: 12-14, 9: 14-16,

10: 16-18, 11:18-20, 12: 20 or more, 13: no answer

B.5. Financial literacy

These questions are from Fernandes, Lynch, and Netemeyer (2014).

1) Imagine that the interest rate on your savings account was 1% per year and inflation was2% per year. After 1 year, would you be able to buy:

- More than today with the money in this account
- Exactly the same as today with the money in this account
- Less than today with the money in this account
- Don't know
- Refuse to answer

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

- True
- False
- Don't know
- Refuse to answer

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

- Savings accounts
- Stocks
- Bonds
- Don't know
- Refuse to answer

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

- Savings accounts
- Stocks
- Bonds
- Don't know

- Refuse to answer

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

- Increase
- Decrease
- Stay the same
- Don't know
- Refuse to answer

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."

- True
- False
- Don't know
- Refuse to answer

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."

- True
- False
- Don't know
- Refuse to answer

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."

- True
- False
- Don't know
- Refuse to answer

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?

- More than $200
- Exactly $200
- Less than $200
- Don't know
- Refuse to answer

10) Which of the following statements is correct?

- Once one invests in a mutual fund, one cannot withdraw the money in the first year
- Mutual funds can invest in several assets, for example, invest in both stocks and bonds
- Mutual funds pay a guaranteed rate of return which depends on their past performance
- None of the above
- Don't know
- Refuse to answer

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

- He owns a part of firm B
- He has lent money to firm B
- He is liable for firm B's debts
- None of the above
- Don't know
- Refuse to answer

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 ittake to eliminate your credit card debt if you made no additional new charges?

- Less than 5 years
- Between 5 and 10 years
- Between 10 and 15 years
- Never
- Don't know
- Refuse to answer

Internet Appendix C (Additional analyses)

Table C1 reports the proportion of males (along with proportion tests), the mean CRT (along with Wilcoxon Rank Sum Tests) for CFAs in the top panel, and students in the bottom panel. About 93% of participants were male across the four tasks for CFAs, and this proportion was 62% for students. The mean CRT was 5.1 for CFAs, and 5.2 for students. The values without parentheses (in parentheses) in the second and fourth columns are the z statistics and p-values for the proportion test for male, and for the Wilcoxon Rank Sum Test for CRT. None of these tests lead to significant differences at a 5% level. Thus, CFAs and students do not significantly differ across the four tasks in terms of CRT scores and gender.

Table C1: Proportion of males and mean CRT (along with tests) for CFAs in the top panel, and students in the bottom panel.

		CFAs		
	Proportion (male)	Proportion test	Mean CRT	Wilcoxon Rank Sum Test
RW	93.84%	*vs* SPdaily 0.230 (0.818) *vs* SPmonthly 0.133 (0.894) *vs* NKI -0.5969 (0.5506)	5.127	*vs* SPdaily 0.153 (0.878) *vs* SPmonthly 0.117 (0.907) *vs* NKI 0.637 (0.524)
SPdaily	93.51%	*vs* SPmonthly -0.098 (0.922) *vs* NKI -0.776 (0.438)	5.113	*vs* SPmonthly -0.043 (0.966) *vs* NKI 0.458 (0.647)
SPmonthly	93.25%	*vs* NKI -0.701 (0.483)	5.098	*vs* NKI 0.511 (0.609)
NKI	95.30%		5.027	
		Students		
	proportion (male)	proportion test	CRT mean	Wilcoxon Rank Sum Test
RW	62.37%	*vs* SPdaily 0.302 (0.762) *vs* SPmonthly -0.015 (0.988) *vs* NKI 0.169 (0.866)	5.297	*vs* SPdaily 0.296 (0.767) *vs* SPmonthly 0.029 (0.977) *vs* NKI -0.125 (0.901)
SPdaily	62.83%	*vs* SPmonthly -0.315 (0.753) *vs* NKI -0.106 (0.915)	5.312	*vs* SPmonthly -0.266 (0.790) *vs* NKI -0.382 (0.702)
SPmonthly	61.24%	*vs* NKI 0.181 (0.856)	5.286	*vs* NKI -0.149 (0.881)
NKI	61.83%		5.307	

Table C2. Multiple Hypothesis Testing

Because we could not conduct multiple hypothesis testing corrections with panel regressions while accounting for dependence across measures, we started by computing the average RFE for each participant in each task. We then applied the technique developed in List, Shaikh and Xu (2019) using linear regressions with CFA Dummy and CRT as controls.[1] Below we provide a summary of the tests along with multiple hypothesis corrections for each task. *P*-values are higher than when using panel regressions (see *t*-test column) as we use only one observation per participant for each task, thus limiting statistical power.

p-values Task Name (Nominally higher RFE: CFA or Elite students)	Standard (*t*-test)	List, Shaikh and Xu (2019)
Task RW (Elite students)	0.949	0.404
Task SPdaily (CFA)	0.028	0.108
Task Monthly (Elite students)	0.044	0.126
Task NKI (CFA)	0.084	0.162

[1] We used the Stata module *mhtreg* (Steinmayr, 2020).

Table C3. Lee bounds estimations (NKI)

Lower / Upper Lee bounds estimates

Dependent variable	Standardized RFE	Standardized Zeal
	(1)	(2)
CFA Dummy	-0.425^{***} / -0.122	-0.301^{***} / -0.065
	(0.067) / (0.069)	(0.064) / (0.067)
Trimming proportion	0.187	0.188

Bootstrapped (5000 reps) standard errors in parentheses.
*** *p*-value < 0.01, ** 0.01< *p*-value < 0.05
(std) stands for standardized.

Table C4. Relative Forecast Error in four tasks
(not controlling for CRT)

Dependent variable	Standardized RFE			
Task	RW (1)	SPmonthly (2)	SPdaily (3)	NKI♠ (4)
CFA Dummy	0.019	-0.037**	0.064**	-0.232**
	(0.075)	(0.017)	(0.029)	(0.100)
Constant	0.224***	-0.015	0.031	0.100
	(0.063)	(0.059)	(0.055)	(0.120)
R^2	0.024	0.164	0.443	0.053
Observations	7,040	7,400	8,138	1,181

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses for regressions, series and order fixed effects included and month for NKI.
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05
CFA Dummy takes value one for participants who are CFAs.
♠ In NKI, we face selection issues as the study takes place over six months, and not all students and CFAs make a forecast in each month. We obtained consistent results when estimating Lee bounds (Lee, 2009) as in Table C4 in Internet Appendix C to correct for selection issues.
(std) stands for standardized.

Table C5. Relative Forecast Error in two treatments of SPmonthly
(not controlling for CRT)

Dependent variable	Standardized RFE	
Task	SPmonthly (Incongruent) (1)	SPmonthly (Congruent) (2)
CFA Dummy	-0.061**	0.00002
	(0.025)	(0.023)
Constant	-0.094	0.003
	(0.078)	(0.081)
R^2	0.202	0.147
Observations	3,660	3,740

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses for regressions, series and order fixed effects included.
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05
(std) stands for standardized.

Table C6. Relative Forecast Error
Only for participants who answered the financial literacy test
(not controlling for either CRT or financial literacy)

Dependent variable	Standardized RFE			
Task	RW (1)	SPmonthly (2)	SPdaily (3)	NKI♠ (4)
CFA Dummy	0.051	-0.045**	0.032	-0.235**
	(0.095)	(0.017)	(0.026)	(0.100)
Constant	0.208***	-0.074	-0.013	0.104
	(0.068)	(0.054)	(0.051)	(0.128)
R^2	0.023	0.161	0.482	0.053
Observations	5,408	6,160	6,578	1,048

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses for regressions, series and order fixed effects included and month for NKI.
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05
CFA Dummy takes value one for participants who are CFAs.
♠ In NKI, we face selection issues as the study takes place over six months, and not all students and CFAs make a forecast in each month. We obtained consistent results when estimating Lee bounds (Lee, 2009) as in Table C3 in Internet Appendix C to correct for selection issues.
(std) stands for standardized.

Table C7. Relative Forecast Error
Only for participants who answered both the financial literacy test and CRT (controlling for CRT)

Dependent variable	Standardized RFE			
Task	RW (1)	SPmonthly (2)	SPdaily (3)	NKI♠ (4)
CFA Dummy	0.035	-0.046**	0.031	-0.238**
	(0.095)	(0.018)	(0.026)	(0.099)
CRT score (std)	-0.091	-0.013	-0.002	0.014
	(0.053)	(0.010)	(0.015)	(0.035)
Constant	0.233***	-0.070	-0.012	0.106
	(0.069)	(0.054)	(0.051)	(0.127)
R^2	0.029	0.161	0.482	0.054
Observations	5,408	6,160	6,578	1,048

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses for regressions, series and order fixed effects included and month for NKI. Only data for which we have both CRT scores and financial literacy
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05
CFA Dummy takes value one for participants who are CFAs.
♠ In NKI, we face selection issues as the study takes place over six months, and not all students and CFAs make a forecast in each month. We obtained consistent results when estimating Lee bounds (Lee, 2009) as in Table C3 in Internet Appendix C to correct for selection issues.
(std) stands for standardized.

Table C8. Relative Forecast Error in four tasks
Only for participants who answered the financial literacy test
(controlling for financial literacy)

Dependent variable	Standardized RFE				
Task	RW (1)	SPmonthly (2)	SPdaily (3)	NKI♠ (4)	NKI♠ (5)
CFA Dummy	0.185 (0.103)	-0.036 (0.019)	0.045 (0.028)	-0.089 (0.103)	-
Financial literacy (std)[2]	-0.133*** (0.047)	-0.009 (0.012)	-0.013 (0.016)	-0.140 (0.098)	-
Financial experience (in years) (std)	-	-	-	-	0.021 (0.021)
Constant	0.197*** (0.068)	-0.075 (0.054)	-0.014 (0.051)	0.057 (0.115)	-0.214*** (0.034)
R^2	0.035	0.161	0.483	0.059	0.097
Observations	5,408	6,160	6,578	1,048	648

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses for regressions, series and order fixed effects included and month for NKI.
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05
CFA Dummy takes value one for participants who are CFAs.
♠ In NKI, we face selection issues as the study takes place over six months, and not all students and CFAs make a forecast in each month. We obtained consistent results when estimating Lee bounds (Lee, 2009) as in Table C3 in Internet Appendix C to correct for selection issues.
(std) stands for standardized.

[2] When putting financial literacy and CFA dummy in the same regression, one could be concerned about multicollinearity issues given their positive correlation (0.503). However, collinearity does not seem to be an issue as the variance inflation factors associated with the two regressors (CFA Dummy and Financial literacy) are below 1.55, regardless of the regression.

Table C9. Relative Forecast Error in two treatments of SPmonthly (controlling for financial literacy and CRT)

Dependent variable	Standardized RFE					
Task		SPmonthly (Incongruent)			SPmonthly (Congruent)	
	(1)	(2)	(3)	(4)	(5)	(6)
CFA Dummy	-0.053**	-0.055**	-0.040	-0.033**	-0.034**	-0.031**
	(0.027)	(0.027)	(0.029)	(0.015)	(0.016)	(0.018)
CRT score (std)	-	-0.011	-	-	-0.021	-
		(0.015)			(0.012)	
Financial literacy (std)[a]	-	-	-0.013	-	-	-0.002
			(0.018)			(0.010)
Constant	-0.141	-0.138	-0.143	-0.062***	-0.057***	-0.062***
	(0.082)	(0.081)	(0.082)	(0.021)	(0.022)	(0.021)
R^2	0.207	0.207	0.207	0.155	0.155	0.155
Observations	3,080	3,080	3,080	3,080	3,080	3,080

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses for regressions, series and order fixed effects included.
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05
(std) stands for standardized.
a: Collinearity does not seem to be an issue as the variance inflation factors associated with the two regressors (CFA Dummy and Financial literacy) are below 1.40, regardless of the regression.

Table C10. Relationship between the Relative Forecast Error in NKI and Forecasting day, Dedication index, and Use of paid services

Dependent variable	Forecasting Day	Standardized RFE		
	(1)	(2)	(3)	(4)
Forecasting Day	-	-0.028 (0.030)	-	-
CFA Dummy	0.315*** (0.076)	-0.223** (0.102)	-0.204 (0.110)	-0.188 (0.106)
CRT score	0.029 (0.039)	-0.0004 (0.024)	-0.029 (0.027)	-0.028 (0.026)
Dedication Index			-0.035 (0.028)	
Use of paid services				-0.029 (0.019)
Constant	23.167*** (0.220)	0.747 (0.724)	0.354 (0.273)	0.272 (0.230)
R^2	0.943	0.054	0.062	0.060
Observations	1,181	1,181	980	980

Linear panel regressions with random effects and robust standard errors clustered at the individual levels in parentheses for regressions, months fixed effects included.
Robust standard errors in parentheses. *** p-value < 0.01, ** 0.01< p-value < 0.05

Table C11. Descriptive statistics across samples and tasks

Task	RW			SPmonthly		
Mean <Median> (Standard Deviation)	CRT	Financial literacy	Risk aversion^{+}	CRT	Financial literacy	Risk aversion
CFA Japan (n = 212, CRT) & (n = 110, Financial literacy)	5.128 <6> (1.227)	10.927 <11> (1.254)	4.142 <4> (1.097)	5.073 <6> (1.395)	11 <11> (1.074)	4.256 <4> (1.004)
E-Student (n = 228)	5.316 <6> (0.899)	8.285 <9> (2.448)	4.263 <4> (1.154)	5.356 <6> (0.832)	8.297 <9> (2.636)	4.188 <4> (1.198)
NE-Student (n = 149)	4.181 <5> (1.586)	7.040 <7> (2.379)	4.409 <4> (1.461)	4.562 <5> (1.435)	6.859 <7> (2.356)	4.578 <4> (1.445)
Wilcoxon Rank Sum Test (p-value)						
CFA vs E-Student	0.396	<0.001	0.160	0.647	<0.001	0.759
CFA vs NE-Student	<0.001	<0.001	0.005	0.008	<0.001	0.044
CFA vs All students	0.009	<0.001	0.021	0.327	<0.001	0.453
E-Student vs NE-Student	<0.001	<0.001	0.099	<0.001	<0.001	0.033

Task	SPdaily			NKI		
Mean <Median> (Standard Deviation)	CRT	Financial literacy	Risk aversion^{+}	CRT	Financial literacy	Risk aversion
CFA Japan (n = 212, CRT) & (n = 110, Financial literacy)	5.113 <6> (1.223)	10.981 <11> (1.127)	4.091 <4> (1.0059)	5.078 <6> (1.316)	11 <11> (1.118)	4.092 <4> (1.060)
E-Student (n = 228)	5.312 <6> (0.903)	8.262 <9> (2.470)	4.262 <4> (1.169)	5.329 <6> (0.941)	8.471 <9> (2.320)	4.219 <4> (1.229)
NE-Student (n = 149)	4.211 <5> (1.557)	7.053 <7> (2.447)	4.391 <4> (1.440)			
Wilcoxon Rank Sum Test (p-value)						
CFA vs E-Student	0.362	<0.001	0.104	0.243	<0.001	0.226
CFA vs NE-Student	<0.001	<0.001	0.004			
CFA vs All students	0.033	<0.001	0.143			
E-Student vs NE-Student	<0.001	<0.001	0.099			

Note: + Number of safe options (Holt and Laury, 2002)