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

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[a], Brice Corgnet[b], Nobuyuki Hanaki[c], Katsuhiko Okada[d], Yohanes E. Riyanto[e], Jiahua Zhu[f]

January 5, 2022

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 tasks. Although students and professionals performed similarly in the artificial forecasting tasks, their performance differed in the more realistic tasks. Yet, increasing the 'representativeness of the situation' in the lab tasks did not systematically benefit financial professionals as students outperformed CFAs when forecasting historical series. However, professionals outperformed students in the field task. Our results imply that the expertise of financial professionals might have been underestimated in previous works that focused on lab tasks.

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) 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 Research Institute of Socionetwork Strategies at Kansai University for their support in conducting our additional experiments, and Yuta Shimodaira for his excellent research assistance. 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).

[a] Division of Economics and NTU-WeBank Joint Research Center, Nanyang Technological University. E-mail: baote@ntu.edu.sg

[b] Emlyon Business School. E-mail: corgnet@em-lyon.com

[c] Corresponding author. Institute of Social and Economic Research, Osaka University. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

[d] Institute of Business and Accounting, Kwansei Gakuin University, and Magne-Max Capital Management. E-mail: katsuokada@magne-max.com

[e] Division of Economics, Nanyang Technological University. E-mail: yeriyanto@ntu.edu.sg

[f] Ma Yinchu School of Economics, Tianjin University. E-mail: jiahuazhu@tju.edu.cn

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 of 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 the roots of success remain secretive. Any enduring recipe for success must indeed remain hidden.

Finance scholars have long argued that no winning recipe can exist because markets are informationally efficient so that all the relevant information for predicting stocks are already incorporated into prices (Samuelson, 1965; Fama, 1965, 1970, 1991; Batchelor, 1990; Hartzmark, 1991; Barber and Odean, 2000; Qu, Timmermann and Zhu, 2019). They postulate that asset prices follow a random walk, and any successful prediction can only be due to luck (Malkiel, 1999; Tetlock, 2009b). At the empirical level, a series of studies have used survey data to evaluate forecasting accuracy following the early work of Cowles (1933). These 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 Philadelphia. Experts, which include US economists from the public and private sectors and academics, are asked to provide stock market forecasts for six-month and twelve-month horizons. The last survey from December 2021 lists twenty-five experts. 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). Anecdotally, Kahneman (2011) also reports, using archival data, that the performance rank of financial advisors is not correlated across years. All these findings echo the work of Tetlock (2009a), who showed the inability of experts to consistently predict major political events and likened forecasting to reading tarot (Tetlock, 2009b).

Despite the early survey results and the numerous claims that there is no expertise in

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

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 (De Bondt and Thaler, 1985, 1987; Lo and MacKinlay, 1988; Bernard and Thomas, 1990; Cutler, Poterba and Summer, 1991; Chopra, Lakonishok and Ritter, 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., 2021) and personality traits such as openness to experience (Gardner and Tetlock, 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, Sasseville, and Williams, 2012; Sprenger et al., 2014).

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, Bennett and Geoum, 1999). Furthermore, the Livingston survey is based only on a handful of experts, which may not represent the population of experts. [3]

Our goal is thus to reexamine the claim that financial experts cannot forecast any better than lay people 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

[2] Philip Tetlock and Daniel Kahneman (see p. 267 of Kahneman, Sibony and Sunstein, 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.

conducted our experiments with financial professionals (Certified Member Analysts of the Securities Analysts Association of Japan) and compared their performance with students often used as a convenience sample in finance experiments (Kirchler et al., 2018).

A few experimental studies have directly compared financial professionals with students, often showing that students perform at least as well as professionals. Glaser, Langer, and Weber (2007) reported no differences in performance in a simple task that consisted in identifying the trend of a stochastic variable. In a herding experiment, Cipriani and Guarino (2009) reported results for financial professionals that were very similar to earlier results 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.[4] 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.

In the studies in which differences were found across samples, professionals often exhibited stronger behavioral biases than students. For example, 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. In particular, professionals were more likely to invest in the stock when they were asked to remember a story stimulating risk-seeking behavior before making their investment decision. Relatedly, Cohn, Fehr, and Maréchal (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, Langer, and Weber (2007) reported higher levels of overconfidence of professionals than students in a laboratory prediction task.

[4] 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).

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). In a few studies, financial professionals have also exhibited fewer biases than students. For example, 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, Haigh, and List (2007) showed that finance professionals did better than students in assessing the precision of signals and were also less likely to be impacted by the presence of losses. However, the authors did not report significant differences in earnings between financial professionals and students (see Result 1, p. 161).

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. The studies comparing students and financial professionals belong to what Harrison and List (2004) defined as 'artefactual field experiments', which are standard lab experiments conducted with non-student subject pools. Previous studies have thus 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:

> "*Third is the representativeness of the situation. A subtle fact is that the research and policy communities oftentimes generalize results to both a population of situations and a population of people, even though we often only speak of the latter. This is particularly troubling considering that the data, thus far, suggest that representativeness of the situation is much more important than the representativeness of the population when it comes to generalizing or scaling (see, e.g., List, 2007).*"

It is interesting to note that, despite focusing on the 'representativeness of the population' by comparing various samples in standard experimental tasks, Snowberg and Yariv (2021, p. 690) also recognized the relevance of the often-neglected ‘representativeness of the situation,’ as highlighted in the following quotation:

> “*For example, we do not address worries that individuals may respond differently to choices that do not mimic the somewhat artificial designs often seen in the lab. We are sympathetic to this view and certainly believe that framing of decisions matters for choices*.”

It follows that financial professionals might perform better in a setting that more closely resembles their daily work environment. Thus, it seems vital to study financial forecasting expertise by comparing professionals and students in various settings, including ‘artefactual field experiments’ and ‘framed field experiments’. The latter involves conducting longitudinal experimental tasks that consist of forecasting actual financial series over a long period of time (Harrison and List, 2004) as in Glaser, Iliewa, and Weber (2019).[5] The latter study 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.

Our study recruits certified financial analysts (CFAs henceforth) from Japan 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. The first two experiments involved simple tasks that consisted of predicting the next value of artificially-created series, as in Bloomfield and Hales (2002) and Glaser et al. (2019) (Tasks RW & SPmonthly, henceforth). The third experiment (Task SPdaily, henceforth) consisted of forecasting historical stock market series. Participants were shown graphs of 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

[5] Beyond forecasting tasks, previous works have also framed their experimental design in a financial context to study bubbles (Kirchler, Lindner, and Weitzel, 2018), risk (Razen, Kirchler and Weitzel, 2020), and investments in structured financial products (Hanaki, 2022).

30th, 2018. They were not told the name of the stock or the starting date of the graph, which was randomly selected. Finally, the fourth experiment (Task NKI, henceforth) asks participants to forecast the closing price of the Nikkei three to four weeks ahead. This task is very similar to the daily jobs of finance professionals in our study and the task used by Glaser et al. (2019) in their Study 3.[6]

Thus, the first three experiments involved lab tasks, whereas the last one featured a field task identical to finance professionals' day-to-day jobs. Although students and professionals performed similarly in the most abstract lab tasks that do not require numeracy skills, their performance differed in the more realistic lab task and in the field task. This shows that increasing the 'representativeness of the situation' affects the relative performance of both samples, in line with List's (2021) claim. However, increasing the 'representativeness of the situation' did not necessarily benefit financial professionals. In the case of our third experiment 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. We interpreted this surprising finding as resulting from experts' excess zeal, where zeal is defined as the difference between a participant's forecast and the last observed value of a financial series. Zealous experts thus made forecasts that sharply differed from the last observed value, which is the best predictor of a stock price if markets are informationally efficient (Fama, 1970). We find that zeal increased substantially when experts were asked to predict an actual series compared to a simulated series, which explains their poor performance in the third experiment. However, professionals outperformed students in the field forecasting task in which their expertise and better access to information could justify bold predictions.

Furthermore, differences in performance across samples were largely mediated by financial literacy, whereas cognitive ability did not play a substantial role. This finding emphasizes that financial forecasting expertise that goes beyond cognitive skills exists. CFAs could afford to be zealous in the field task (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 the relationship between the 'representativeness of the situation' and the relative performance of experts is not

[6] This type of forecasting task also captures the essence of Keynes' "beauty contest" metaphor of financial markets (Nagel, 1995, Hommes, 2021).

monotonic. As a result, it might be difficult to assess the true extent of financial professionals' expertise in lab studies. The expertise of financial professionals might thus have been underestimated in previous works that focused on lab tasks. Yet, in line with Fréchette (2015), we should acknowledge that the size of the difference in performance between students and professionals remains small, even in the field task (Cohen's $d = 0.243$).

2. Design

2.1. Participants

We recruited financial professionals who were CFAs and students from Osaka University. Professionals were certified members of the Securities Analysts Association of Japan (SAAJ) that were recruited thanks to the association's support. We were able to invite more than 16,000 members of SAAJ to participate in our experiment, of which 255 initially registered for the experiment.

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 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 acute 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, Lynch, and Netemeyer, 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 participants.

Our university participants were enrolled at Osaka University, widely considered an elite university in Japan. According to entry exams 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 in the Top 5 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.

2.2. Protocol

Participants were involved in a longitudinal study that consisted of 4 forecasting tasks. The tasks were as follows:

a) **RW**: **Forecasting random walk time series**. We asked participants to predict the pre-generated (artificial) time series~~.~~

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 stock's name and the selected historical sample. Participants received graphical information of monthly values of either prices or returns.

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 the instructions for each task when recruited, they were told that the last task consisted of forecasting the Nikkei index. Each task was completed sequentially. 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. 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. 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 for CFAs and ¥20 for students. The payment scheme for NKI was based on a tournament structure where the best CFA (student) performer earned ¥5,000 (¥1,000).

In all tasks, the payments offered to CFAs were five times greater than for students to compensate the difference in hourly wages for these two populations, considering the hourly wage of undergraduate students at the university is about ¥950 and the average hourly wage of CFAs in Japan being about ¥4500.[7] 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.

2.3. Tasks

2.3.1. RW: Predicting Random Walk Time Series

This task was based on the experimental design of Bloomfield and Hales (2002). 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:

> *"We have constructed a model of a random process that works much like flipping a faircoin. 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 Appendix A.1 for graphs and detailed instructions).[8]

[7] 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. With standard working hours (8 hours/day, 20 days/months), it is approximately ¥4300.
[8] The English translation of the instructions can also be found at: https://bgt.au1.qualtrics.com/jfe/form/SV_9RAxWjjGFww0zSl.

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 [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.[9]

212 professionals (CFAs) and 228 students participated in the task, lasting around 20 minutes.[10] There were 93.4% male participants among CFAs. These professionals had an average of more than 10 years of experience and an annual gross income greater than ¥10M ($100,000). There were 53.5% male students in our sample, resulting in a relatively balanced sample across the two genders among our student participants. CFAs earned an average of ¥1,362, and the maximum payment was ¥11,000.[11] Students earned an average of ¥316, and the maximum payment was ¥3160.[12]

In addition, we elicited participants' risk attitude (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 (Frederick, 2005), and basic demographics (see Appendix B for details on these additional measures). We also administered a financial literacy quiz (Fernandes, Lynch, and Netemeyer, 2014).

[9] 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.

[10] 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.

[11] The average payment for those who have been selected for the additional reward was ¥6,120.

[12] The average payment for those who have been selected for the additional reward was ¥1,524.

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

We used the 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, where the first dimension varied 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. This procedure was such that the generated line charts did not resemble the graphical information professionals typically face in their daily job (see Appendix A.2 for graphs and detailed instructions).[13-14] We refer to these charts as *unnatural-display* series.

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:

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

[13] The English translation of the instructions can also be found at: https://bgt.au1.qualtrics.com/jfe/form/SV_74H7IiRiVqUQCzA.

[14] 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.

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:

$$max\,\{200 - 10 \times |return\,forecast - realized\,return\ \times 100|, 0\}$$

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.

174 professionals (CFAs) and 206 students participated in this task, which lasted about 20 minutes.[15] Each professional (student) was assigned at random to one of the four treatments. CFAs earned an average of ¥1,675, and the maximum payment was ¥10,930.[16] Students earned an average of ¥284, and the maximum payment was ¥2,170.[17]

We referred to the treatments in which the information provided on the chart (price or returns) and the variable to be forecasted the same as ***congruent***. By contract, ***incongruent*** treatments were those in which the information provided on the chart differed from the variable to be forecasted.

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

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

[15] 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.
[16] The average payment for those who have been selected for the additional reward was ¥8338.
[17] The average payment for those who have been selected for the additional reward was ¥1441.

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. 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 Appendix A.3 for graphs and detailed instructions), as is shown in Figure 1.[18] We refer to these charts as *natural-display* series.

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

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

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

198 professionals (CFAs) and 221 students participated in this task, which lasted about 20 minutes.[19] CFAs earned an average of ¥1,840, and the maximum payment was ¥10,980.[20] Students earned an average of ¥315, and the maximum payment was ¥2,120.[21]

2.3.4. NKI: Forecasting the Nikkei index

[TABLE 2 AROUND HERE]

[18] The English translation of the instructions can also be found at: https://bgt.au1.qualtrics.com/jfe/form/SV_0pmGV8LB9DMRPc9.
[19] 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.
[20] The average payment for those who were selected for the additional reward was ¥8,233.
[21] The average payment for those who were selected for the additional reward was ¥1,694.

In this task, we asked participants to predict the Nikkei index once a month for several months.[22-23] 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 appeal to 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.

As for the previous tasks, the payoff depended on forecasting accuracy. 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), there was no participation fee for this task. In case of a tie, one winner was chosen randomly among the top performers (see Appendix A.4 for detailed instructions).[24-25]

After each forecasting day, participants received graphical feedback, separately for students and CFAs, showing the realized closing Nikkei index along with their forecast 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 Appendix A.4.2).

3. Results

3.1. Forecasting performance of CFAs and students

[22] 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.
[23] 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.
[24] Ties happened in Wave 1 and 3 for students (2 participants) and in Wave 3 for CFAs (3 participants).
[25] The English translation of the instructions can also be found at: https://bgt.au1.qualtrics.com/jfe/form/SV_1MNumMb42kKnpDo

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).[26] 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]

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 (Cohen's d = 0.025). 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 understand better the concepts of probability and randomness (see, e.g., Toplak, West and Stanovich, 2011, 2014).

The SPmonthly task requires financial skills related to numeracy (Glaser, Iliewa, and Weber, 2019) that are especially relevant to financial professionals and a key component of financial literacy (see, e.g., Fernandes, Lynch, and Netemeyer, 2014). Thus, it is not surprising that CFAs outperformed students in SPmonthly (see regression (2)). However, the effect size is relatively small (Cohen's d = 0.278).

[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)). In that case, CFAs might have performed better than students because they possess greater numeracy skills (see financial literacy scores in Table 6), which are required to convert prices to returns and vice versa.

[26] Similarly, Corgnet et al., (2021) use *price forecasting error* as their accuracy measure, which is defined as RFE×value.

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 analysis tools. In that context, CFAs could not outperform students by using professional tools or their knowledge they could exploit better than students. Interestingly, CFAs underperformed students (see regression (3)), although the size of the effect was small (Cohen's d = 0.156). In contrast, CFAs outperformed students (see regression (4)) (Cohen's d = 0.243) in the field task (the 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.[27]

3.2. On zeal and forecasting strategies

If markets are informationally efficient (Fama, 1970), then the forecast that minimizes the RFE equals the last observed value of a financial series. According to that view, we should interpret any forecast deviating from the last observed value of the stock as financial *zeal*. We use the term *zeal* to refer to the urge of financial professionals to show off their expertise by providing detailed forecasts that purposefully differ from the last observed value.[28] Zeal relates to overconfidence (e.g., Hirshleifer and Subrhamanyam, 1998; Barber and Odean, 2001; Gervais and Odean, 2001) because professionals might state bold forecasts when they believe their private information is more precise than it is (e.g., Alpert and Raiffa, 1982) or when they hold optimistic views of the future (Weinstein, 1980). We use the term zeal to refer to a situation in which financial professionals might exhibit overconfidence or optimism to demonstrate their expertise.

[27] 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.

[28] According to Merriam-Webster, zeal is "a strong feeling of interest and enthusiasm that makes someone very eager or determined to do something."

Formally, we define zeal as follows: $100 \times \frac{|forecast - last\ value|}{last\ value}$. Where the last value is the last observed stock price. Zeal is thus not defined for the artificial series used in RW. In SPmonthly, the measure of zeal tends to be larger for *incongruent* than for *congruent* treatments. As a result, any discrepancy between the last observed value of a stock and a participant's forecast is likely driven by the type of series used in the forecasting task rather than zealous behavior.[29]

We will thus focus on zeal as a potential explanation for the difference in forecasting performance between CFAs and students in SPdaily and NKI. In NKI, the last observed value of the Nikkei index is the last closing price of the index participants could have observed before submitting their forecast.[30]

[TABLE 5 AROUND HERE]

In Table 5, we show that CFAs tend to be more zealous than students in SPdaily, while this is not the case in NKI, thus suggesting that field tasks led them to be more cautious than students by providing forecasts that were closer to the last observed value of stocks. Zeal leads to worse forecasting performance (see regressions (3) and (4) in Table 5) in both SPdaily and NKI for both CFAs and students. However, the negative impact of zeal is less pronounced for CFAs in NKI. This could be because bold forecasts might result, unlike the SPdaily task, from specific financial knowledge of the Nikkei index.

This could also have been partly explained by CFAs being more dedicated to the task and thus 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 C3 in Appendix C), controlling for the date at which a forecast was made did not alter the findings reported in regression (4) of Table 3 (see regression (2) in Table C3 in Appendix C).

[29] See Table C1 in Appendix C.

[30] Since the Nikkei closes at 3pm, we set 3pm as the cut-off time in determining the closing price. Namely, if participants submitted their forecast at 3pm or earlier on a given day, the last observed closing price was set at the previous business day closing value, while for participants who submitted their forecast after 3pm on a given day, the last observed closing price was set at the current day closing value.

In the same vein, we controlled for the level of interest and effort (see Intrinsic Motivation Inventory scale, Ryan, 1982) participants put into the task to assess whether the superior performance of CFAs was simply due to their greater dedication to the task. This effort and dedication index (Cronbach's α= 0.783) was, however, not significantly different between CFAs (3.320) and students (3.081) (Rank Sum Test = 0.143). Importantly, controlling for the dedication index, the CFA Dummy in Table 4 (regression (4)) continues to be significant (see regression (3) in Table C3 in Appendix C) although only marginally (p-value = 0.063). The magnitude of the coefficient for CFA Dummy only decreases by 12.1% when controlling for the dedication index, thus suggesting that the mediating role of dedication is limited. 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 C3) as the coefficient for CFA Dummy only decreases by 15.70% when controlling for paid services.

3.3. Financial literacy, cognitive skills, and forecasting performance

This section aims to uncover the mechanisms by which CFAs' forecasting performance differs from students' performance.

3.3.1. 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, 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 students obtained CRT scores as high as 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 over all waves.[31] Not surprisingly, professionals scored substantially higher on financial literacy than students (p-value < 0.001). We then assessed the mediating role of financial literacy in predicting relative forecast error using the same procedure as for CRT (see Tables C6, C8, and C9).[32] 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 coefficient associated with CFAs dummy becomes statistically non-significant when we introduce financial literacy in the regression (see regression (3) in Table C9 and regression (4) Table C8, respectively). The magnitude of the coefficient decreases by 27.01% and 62.18% for the incongruent treatments of the SPmonthly and the NKI, respectively. In sum, although CFAs outperformed students when controlling for CRT, no differences were found 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, Roider and Schmitz, 2009; Pennycook et al., 2012; Shenhav, Rand and Greene, 2012; Weller et al., 2013; Campitelli and Gerrans, 2014).

3.3.2. 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, DeSantis and Porter, 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-

---

[31] 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.
[32] 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.

Students, henceforth) for the first three forecasting tasks in September 2021.[33] In Table 6, we show that NE-Students score significantly lower than E-Students on cognitive ability and financial literacy (p-values < 0.001). 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 on financial literacy.

[TABLE 6 AROUND HERE]

Figure 3 shows that NE-Students tend to produce larger forecasting errors than E-Students and CFAs.

[FIGURE 3 AROUND HERE]

In Table 7, we find that NE-Students performed significantly worse than E-Students and CFAs in RW and SPmonthly. 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. The corresponding p-values are 0.001, 0.016 and 0.130 in regressions (1), (2), and (3) compared to 0.011, 0.096, and 0.739 in regressions (4), (5), and (6). This 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.

Table 7 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. This analysis is shown in Table 8. 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 for the sample of participants who answered both the financial

[33] Because these experiments were conducted in September 2021, we could not gather data for the NKI for this additional sample.

literacy and CRT questions (see regressions (2), (6) and (8)). 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 7 AROUND HERE]

[TABLE 8 AROUND HERE]

4. 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 financial professionals could outperform elite students on various financial forecasting tasks requiring numeracy skills (SPmonthly, incongruent treatments) or using actual financial series in a field setting (the NKI). Cognitive skills like CRT scores did not drive the result because the level of cognitive ability is similar in the two groups. Thus, our findings can serve as evidence of financial professionals' special skills that go beyond IQ, luck, and the alleged "bragging" culture in the financial industry.

Increasing the 'representativeness of the situation' (List, 2021) impacted our assessment of the relative forecasting performance of CFAs, but the relationship was not monotonic. Although the two groups did not differ in their forecasting performance on the artificial forecasting task (RW), elite students outperformed CFAs in forecasting historical financial series in a controlled setting.

Our study has three main implications. First, the extent of financial professionals' skills cannot be reliably measured using only lab tasks. Second, increasing the 'representativeness of the situation' will not necessarily favor financial professionals because applying financial knowledge to a forecasting task that looks superficially similar to professionals' daily jobs might lead to excess zeal. We found excess zeal and poor performance when professionals had to forecast historical financial series in a controlled setting without any information but the graphs of the series. Third, we showed evidence that CFAs possess forecasting skills to outperform elite students. Furthermore, the extent to which CFAs outperformed students was mediated by financial literacy but not by CRT scores, thus suggesting financial knowledge rather than cognitive skills explains performance in financial forecasting tasks.

References

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

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," *The Journal of Finance*, 59(3), 1259-1294.

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

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

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.

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.

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

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.

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

Corgnet, B., Deck, C., Desantis, M., & Porter, D. (2020) "Forecasting Skills in Experimental Markets: Illusion or Reality?" Available at SSRN 3645967.

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

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

De Bondt, W. P. (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.

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

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.

Gervais, S. & 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

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.

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.

Hommes, C.H., (2021). “Behavioral and experimental macroeconomics and policy analysis: a complex systems approach,” *Journal of Economic Literature*, 59(1), 149-219.

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

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,” *The Journal of Finance*, 35(4), 921-931.

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

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

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. (2021) “The Voltage Effect in Behavioral Economics” (No. 00733). The Field Experiments Website.

List, J. A. (2003) “Does market experience eliminate market anomalies?” *The*

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.

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," *The 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.

Nagel, R., (1995). "Unraveling in guessing games: an experimental study," *American Economic Review*, 85(5), 1313-1326

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.

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.

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.

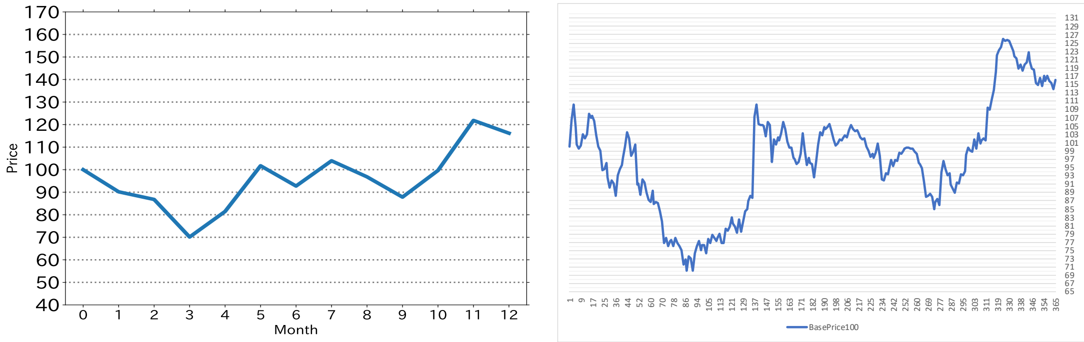
Toplak, M. E., West, R. F., & Stanovich, K. E. (2014) “Assessing miserly information processing: An expansion of the Cognitive Reflection Test,” *Thinking & Reasoning*, 20(2), 147-168.

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

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).

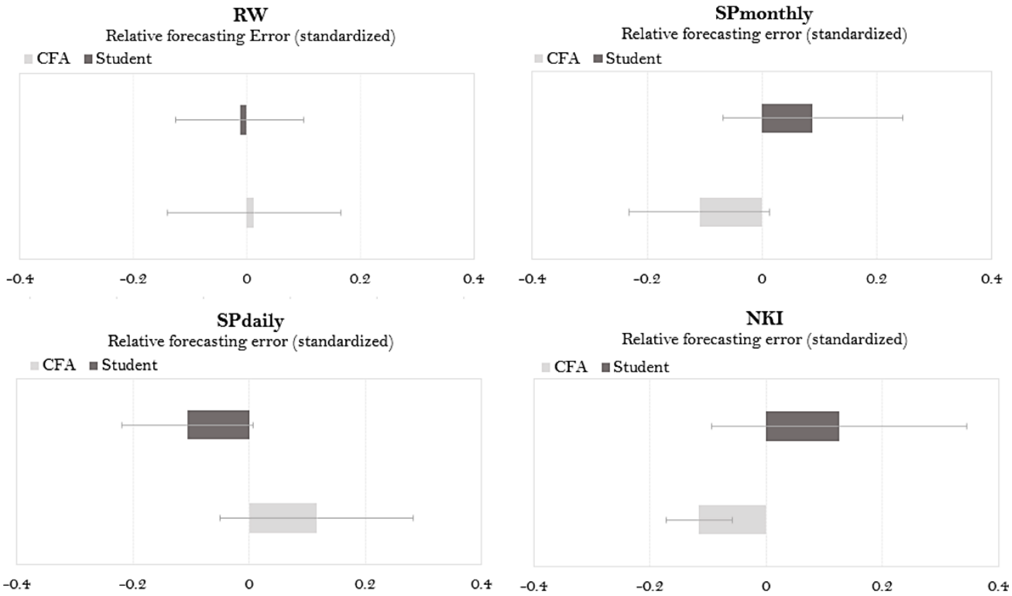


Figure 2. Average standardized RFE (computed at the participant level) for each task across CFAs and students along with 95% confidence intervals.

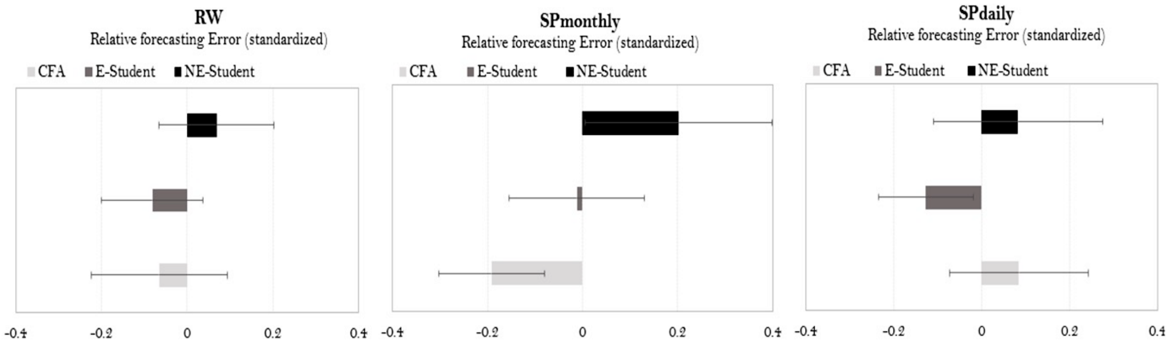


Figure 3. Average standardized RFE (computed at the participant level) for each task for CFAs and the two student samples along with 95% confidence intervals.

Tables

Table 1. Description of forecasting tasks

Tasks	Series	Display	Information	Nature of the task	Time frame	Adapted from	Number of participants	
							CFAs	Students
RW	Artificial	Unnatural	Series	Lab	Static	Bloomfield and Hales (2012)	212	228
SPmonthly (congruent)	Real	Unnatural	Series	Lab	Static	Glaser, Iliewa and Glaser (2019) [Studies 1 & 2]	87	105
SPmonthly (incongruent)	Real	Unnatural	Series (incongruent)	Lab	Static		87	101
SPdaily	Real	Natural	Series	Lab	Static	This paper	198	221
NKI	Real	Natural	Complete	Field	Longitudinal	Glaser, Iliewa and Glaser (2019) [Study 3]	717 (Cumulative Total. See Table 2 for the breakdown)	538
		Cumulative Total Number of Participants					1,301	1,193

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 Participants		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. Relative Forecast Error

Dependent variable	Standardized RFE			
Task	RW (1)	SPmonthly (2)	SPdaily (3)	NKI♠ (4)
CFA Dummy (std)	0.002	-0.018**	0.026**	-0.116**
	(0.037)	(0.008)	(0.013)	(0.050)
CRT score (std)	-0.099**	-0.011	-0.059*	0.0006
	(0.045)	(0.009)	(0.034)	(0.028)
Constant	0.253***	-0.025	0.065	-0.020
	(0.057)	(0.059)	(0.055)	(0.080)
R^2	0.030	0.164	0.445	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, ** *p*-value < 0.05, * *p*-value < 0.1
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 thus estimated Lee bounds (Lee, 2009) in Table C2 in Appendix C (see regression (1)) to correct for selection issues. Both the lower and upper bounds of the CFA Dummy coefficient are negative. This might not be surprising given that the level of participation is relatively high on average (64.2%).
(std) stands for standardized.

Table 4. Relative Forecast Error (SPmonthly)

Dependent variable	Standardized RFE	
Task	SPmonthly (Incongruent) (1)	SPmonthly (Congruent) (2)
CFA Dummy (std)	-0.029**	-0.001
	(0.012)	(0.011)
CRT score (std)	-0.006	-0.024
	(0.011)	(0.015)
Constant	-0.112	0.007
	(0.077)	(0.084)
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, ** *p*-value < 0.05, * *p*-value < 0.1
(std) stands for standardized.

Table 5. Zeal

Dependent variable	Standardized Zeal		Standardized RFE	
Task	SPdaily (1)	NKI♣ (2)	SPdaily (3)	NKI (4)
Standardized Zeal	-	-	1.070*** (0.066)	0.923*** (0.022)
CFA Dummy (std)	0.029** (0.012)	-0.070 (0.050)	-0.005 (0.006)	-0.052*** (0.0321)
CFA Dummy (std) × Standardized Zeal	-	-	0.015 (0.12051)	-0.055*** (0.027)
CRT score (std)	-0.051 (0.032)	-0.009 (0.028)	-0.003 (0.007)	-0.006 (0.008)
Constant	-0.116*** (0.033)	0.138* (0.078)	-0.060 (0.044)	-0.161*** (0.009)
R^2	0.063	0.007	0.759	0.924
Observations	8,138	1,181	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, ** p-value < 0.05, * p-value < 0.1
♣ 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 thus estimated Lee bounds (Lee, 2009) in Table C2 in Appendix C (see regression (2)) to correct for selection issues. Both the lower and upper bounds of the CFA Dummy coefficient are negative. This might not be surprising given that the level of participation is relatively high on average (64.2%).
(std) stands for standardized.

Table 6. CRT and financial literacy

	CRT	Financial literacy
CFA		
Japan (*n*=212, CRT) & (*n*=110, Financial literacy)	5.127	10.927
	(1.227)	(1.254)
E-Student (*n*=228)	5.316	8.285
	(0.899)	(2.448)
NE-Student (*n*=149)	4.181	7.040
	(1.586)	(2.379)
Wilcoxon Rank Sum Test (*p*-value)		
CFA vs E-Student	0.396	<0.001
CFA vs NE-Student	<0.001	<0.001
CFA vs All students	0.009	<0.001
E-Student vs NE-Student	<0.001	<0.001

Table 7. Relative Forecast Error and CRT as a mediator

Dependent variable	Standardized RFE					
Task	RW		SPmonthly		SPdaily	
	(1)	(2)	(3)	(4)	(5)	(6)
CFA Dummy (std)	0.009	0.006	-0.017**	-0.018**	0.030**	0.027**
	(0.036)	(0.036)	(0.008)	(0.008)	(0.014)	(0.013)
NE Student Dummy (std)	0.100***	0.084**	0.023**	0.017*	0.021	0.005
	(0.029)	(0.033)	(0.010)	(0.010)	(0.014)	(0.016)
CRT score (std)	-	-0.043	-	-0.016**	-	-0.042*
		(0.031)		(0.008)		(0.022)
Constant	0.259***	0.259***	0.021	0.019	0.040	0.040
	(0.050)	(0.050)	(0.055)	(0.055)	(0.046)	(0.046)
CFA vs NE Student (*p*-value)	0.014	0.044	<0.001	0.0003	0.557	0.255
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 for regressions, series and order fixed effects included.
Robust standard errors in parentheses. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.1
NE-Student Dummy takes value one for participants who belong to the non-elite university.
(std) stands for standardized.

Table 8. Relative Forecast Error and financial literacy as a mediator

Dependent variable	Standardized RFE											
Task	RW				SPmonthly				SPdaily			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(7)	(8)	(9)	(10)
CFA Dummy (std)	0.025	0.022	0.062	0.059	-0.020**	-0.021**	-0.016*	-0.020**	0.015	0.014	0.020**	0.018**
	(0.046)	(0.046)	(0.049)	(0.807)	(0.008)	(0.008)	(0.009)	(0.009)	(0.012)	(0.012)	(0.014)	(0.009)
NE Student	0.100***	0.089***	0.084***	0.081***	0.023**	0.016*	0.022**	0.016**	0.021	0.017	0.019	0.016**
Dummy (std)	(0.029)	(0.033)	(0.031)	(0.034)	(0.009)	(0.010)	(0.010)	(0.011)	(0.014)	(0.014)	(0.015)	(0.008)
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)		(0.008)
Financial[a] literacy	-	-	-0.076**	-0.073*	-	-	-0.007	-0.002	-	-	-0.010	-0.007
(std)			(0.036)	(0.037)			(0.010)	(0.010)			(0.017)	(0.009)
Constant	0.254***	0.259***	0.266***	0.266***	-0.017	-0.019	-0.015	-0.018	-0.007	-0.007	-0.006	-0.006
	(0.058)	(0.050)	(0.058)	(0.058)	(0.054)	(0.054)	(0.054)	(0.053)	(0.042)	(0.042)	(0.043)	(0.044)
CFA vs NE Student (*p*-value)	0.105	0.164	0.679	0.693	<0.0001	0.0002	0.002	0.003	0.708	0.866	0.958	0.922
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 for regressions, series and order fixed effects included.
Robust standard errors in parentheses. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1
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 and Financial literacy) are below 1.45, regardless of the regression.

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.

Example 1 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.

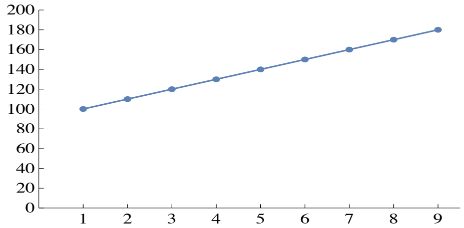
Example 2 If you set the price for the bet to be 50 points and the randomly generated number is 60, you receive 60 points.

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

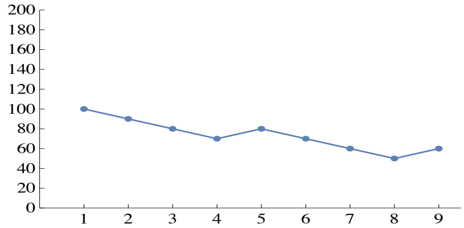
A.1.2. Figures

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.

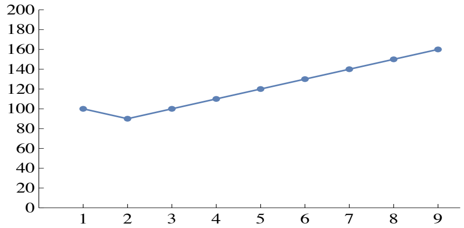
Series 1



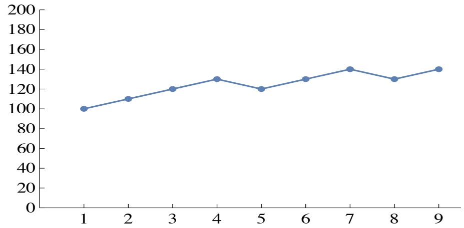
Series 5



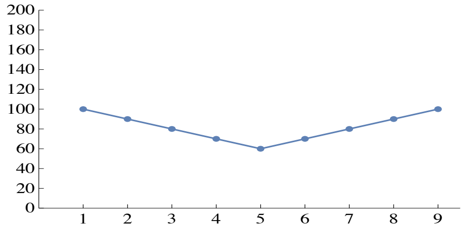
Series 2



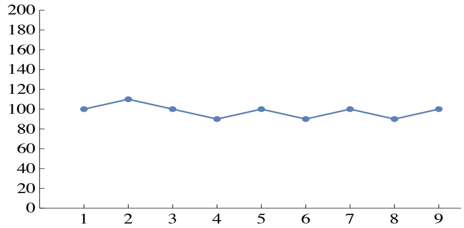
Series 6



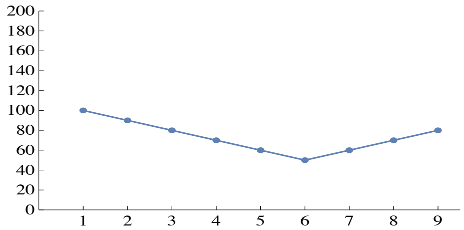
Series 3



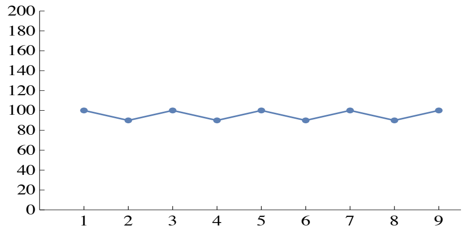
Series 7



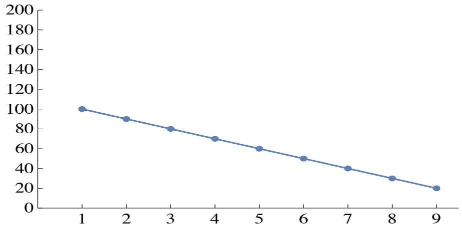
Series 4



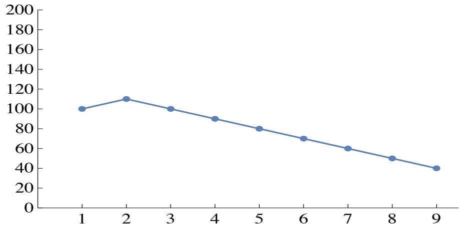
Series 8



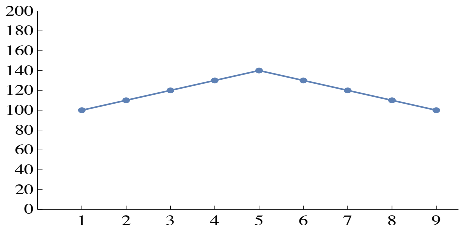
Series 9



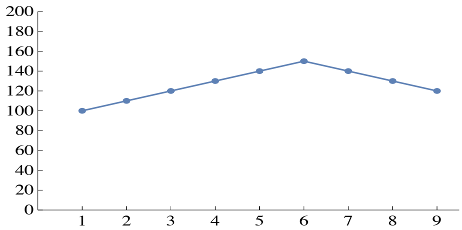
Series 10



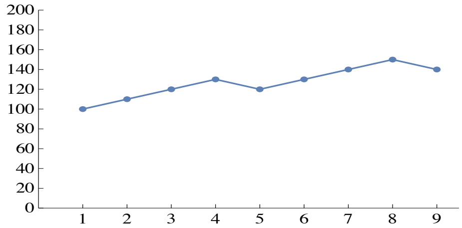
Series 11



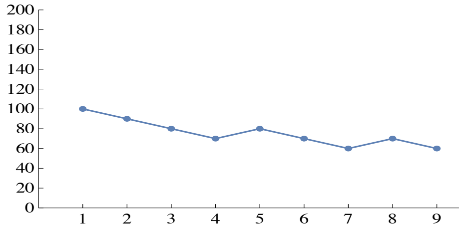
Series 12



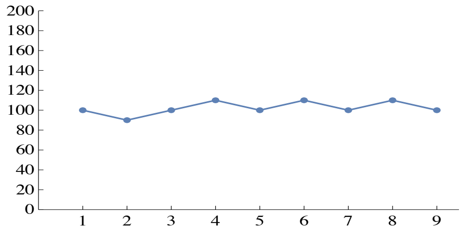
Series 13



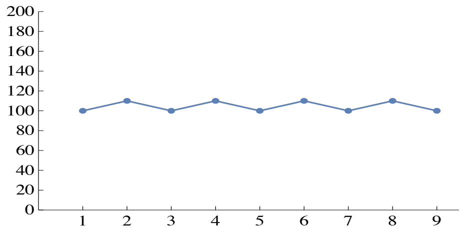
Series 14



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 1 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.

$$max\ \{200 - 10 \times |return\ forecast - realized\ return\ \times 100|, 0\}$$

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.

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.

<u>Incongruent treatment with price predictions</u>

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.

$$max\,\{200 - 10 \times |return\ forecast - realized\ return\ \times 100|, 0\}$$

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.**

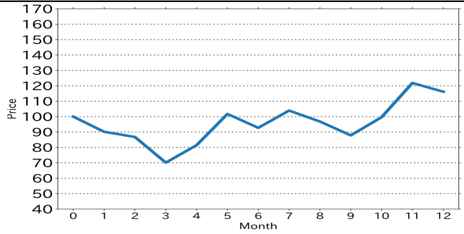
You will do a similar forecasting task for all the 20 randomly chosen stocks. One in ten participants will be selected for financial reward. If you are selected for the financial reward, one of the 20 predictions will be selected randomly, and you will receive the reward according to the points you have 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

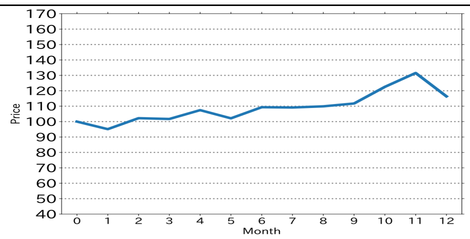
Price series

Series 1

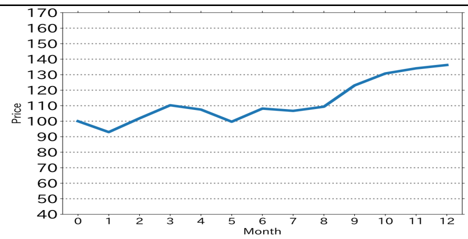


Series 2

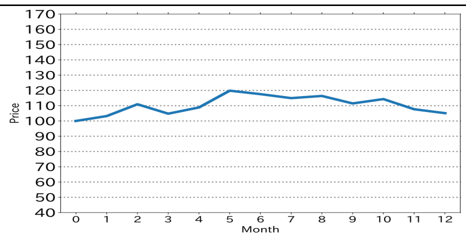
Series 3



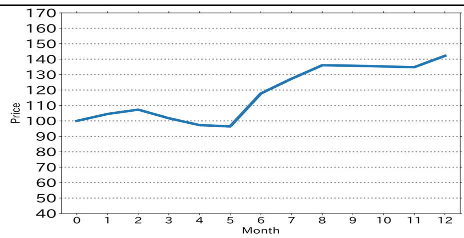
Series 4



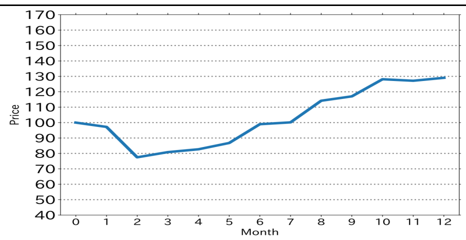
Series 5



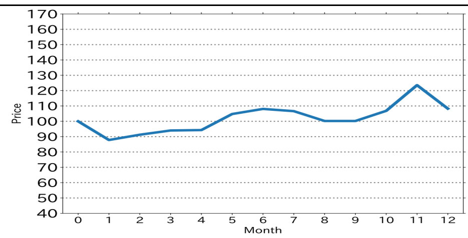
Series 6



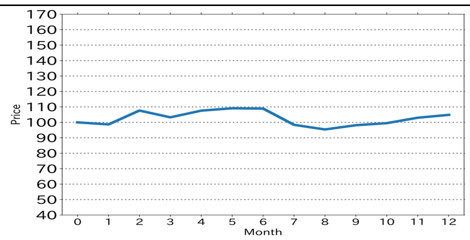
Series 7



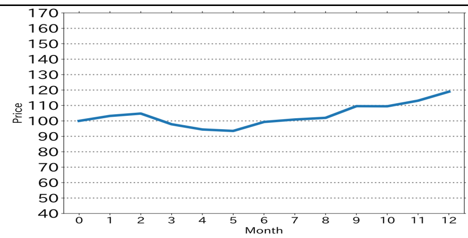
Series 8



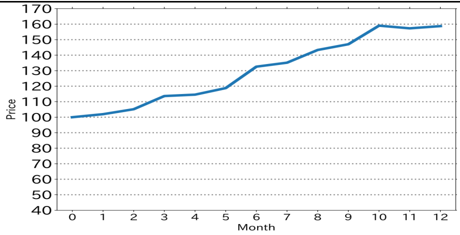
Series 9



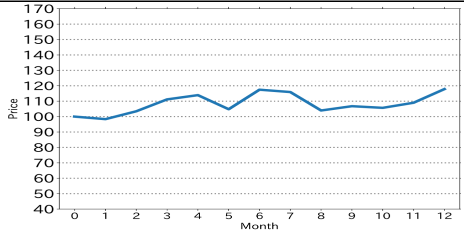
Series 10



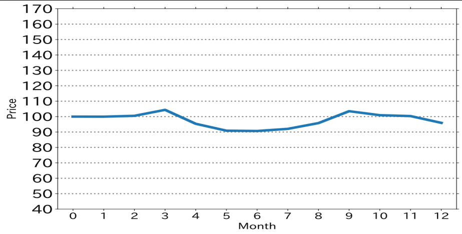
Series 11



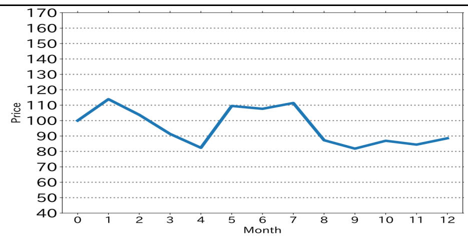
Series 12



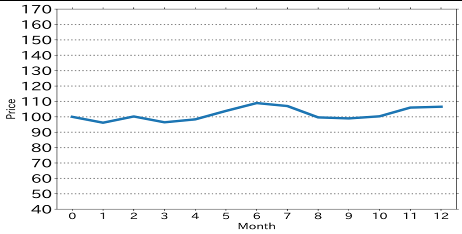
Series 13



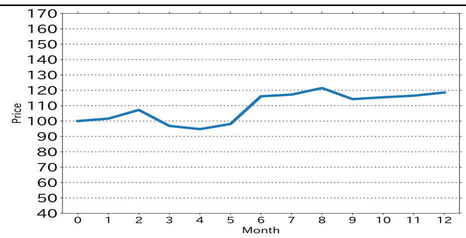
Series 14



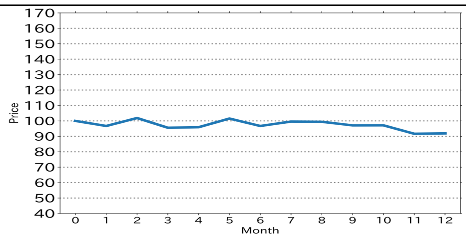
Series 15



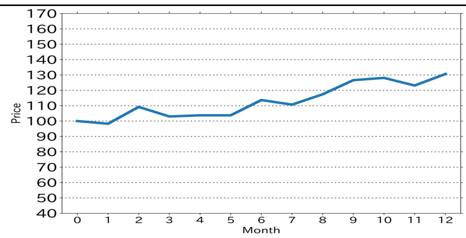
Series 16



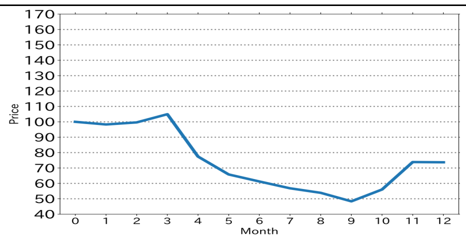
Series 17



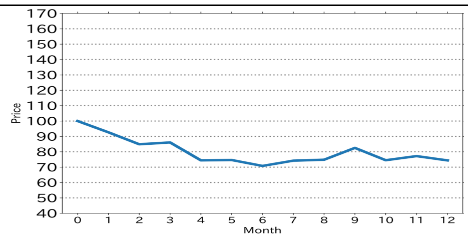
Series 18



Series 19

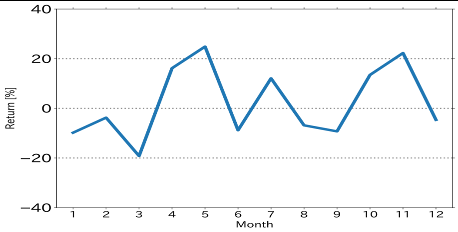


Series 20

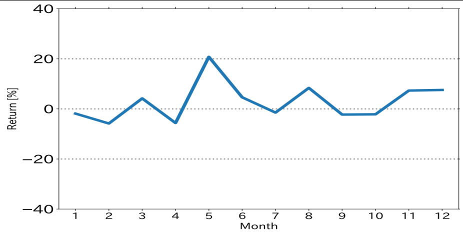


Returns series

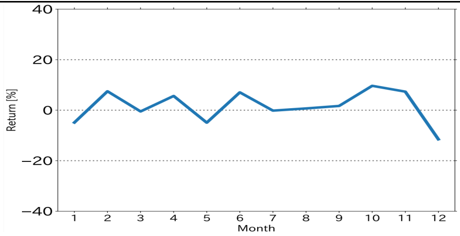
Series 1



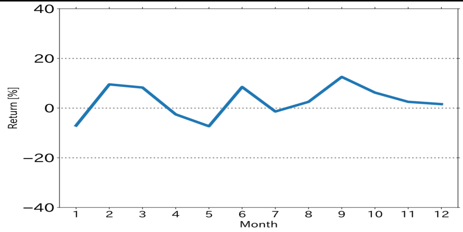
Series 2



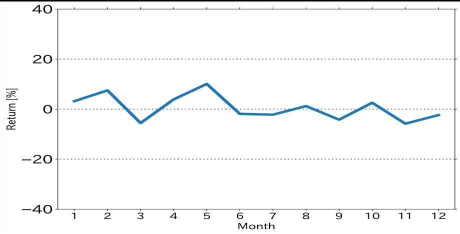
Series 3



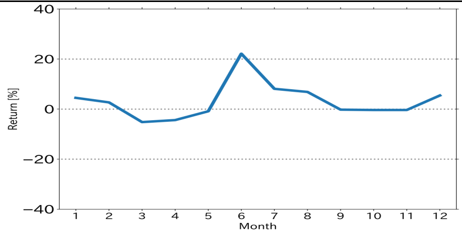
Series 4



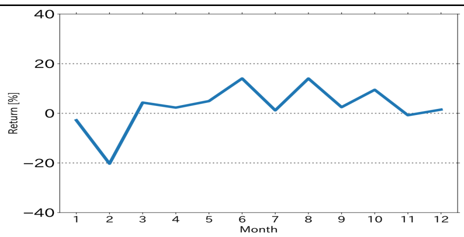
Series 5



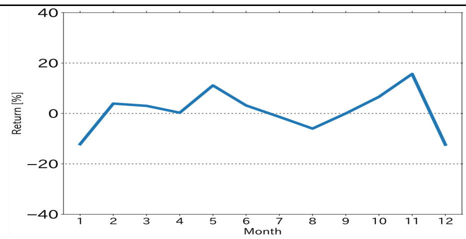
Series 6



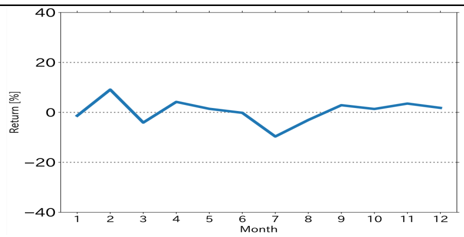
Series 7



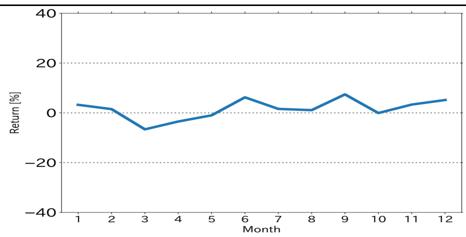
Series 8



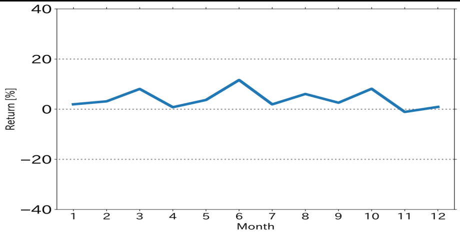
Series 9



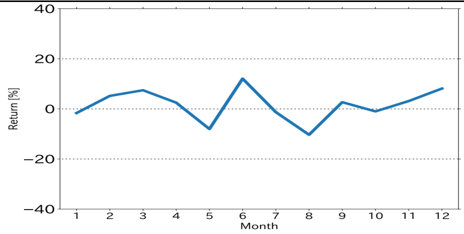
Series 10



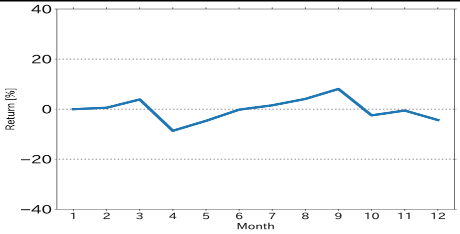
Series 11



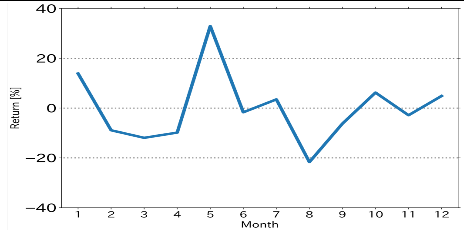
Series 12



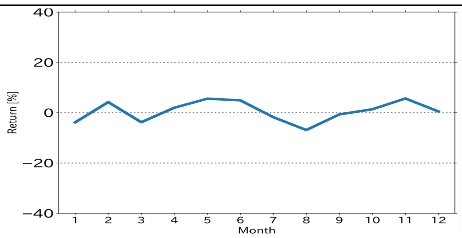
Series 13



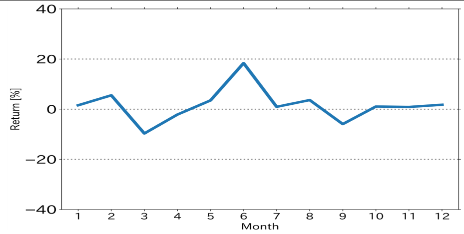
Series 14



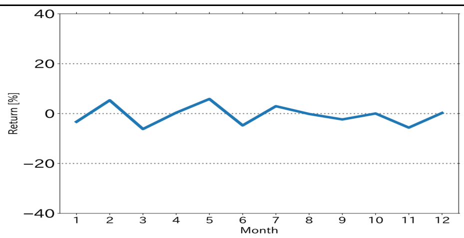
Series 15



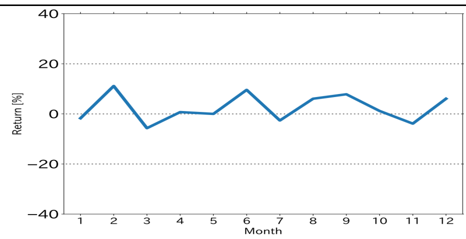
Series 16



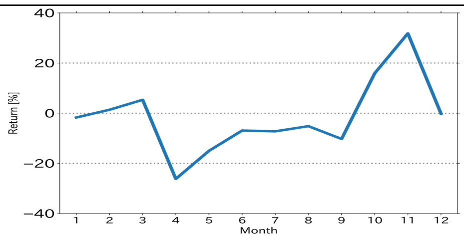
Series 17



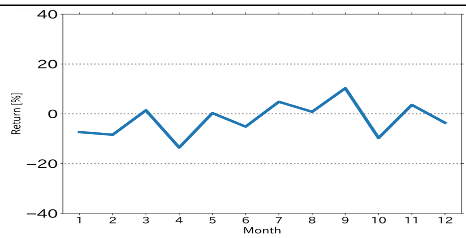
Series 18



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.

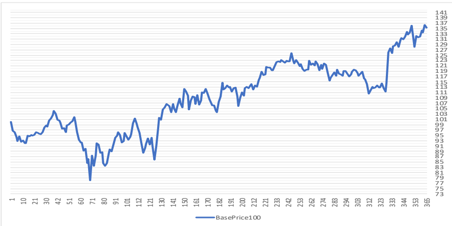
You will not be informed about the accuracy of your forecast until the end of the experiment

A.3.2. Figures

Series used in SPdaily.

Series 1

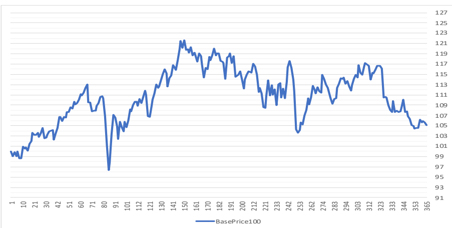
Series 2



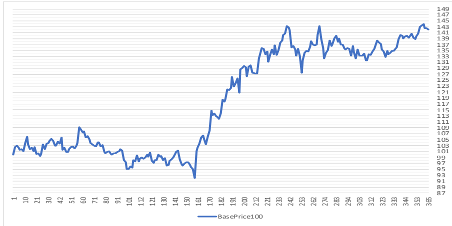
Series 3

Series 4

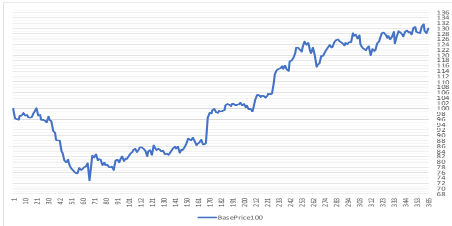
Series 5



Series 6



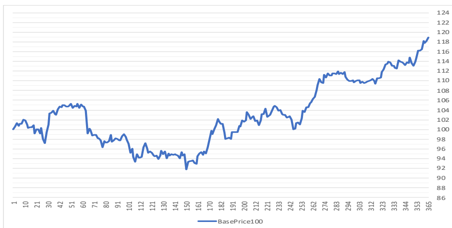
Series 7



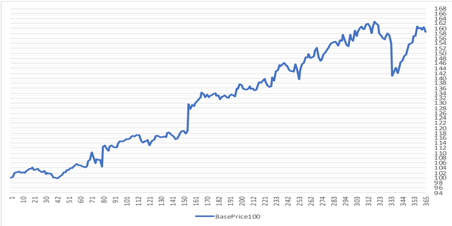
Series 8

Series 9

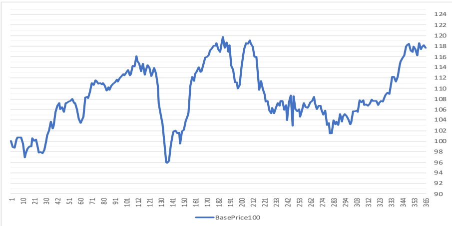
Series 10



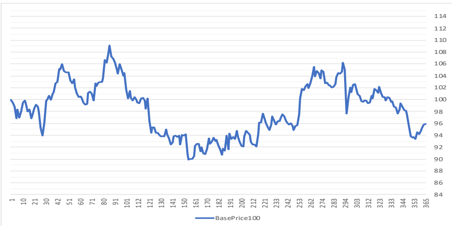
Series 11



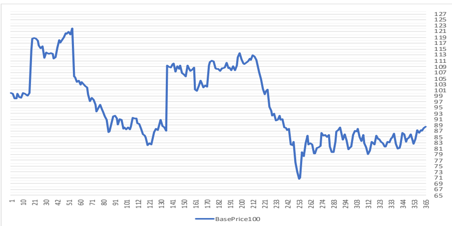
Series 12



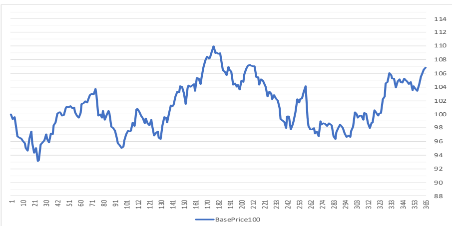
Series 13



Series 14

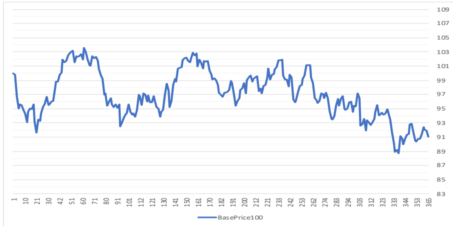


Series 15



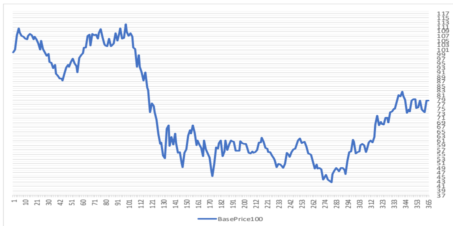
Series 16

Series 17



Series 18

Series 19



Series 20

A.4. NKI.
A.4.1. Forecasting task instructions

Example of feedback and instruction screens.

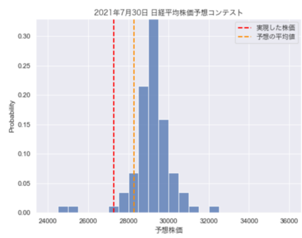
Result of the previous experiment.

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.

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 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

Appendix C (Additional analyses)

Table C1. Zeal (SPmonthly)

Dependent variable	Standardized zeal
Incongruent Dummy (std)	0.108*** (0.023)
CFA Dummy (std)	-0.038* (0.020)
CFA Dummy (std) × Incongruent Dummy (std)	-0.041** (0.020)
CRT score (std)	-0.022 (0.022)
Constant	0.011 (0.028)
R^2	0.062
Observations	7,400

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, ** *p*-value < 0.05, * *p*-value < 0.1
Incongruent Dummy takes value one for incongruent treatments in SPmonthly .
(std) stands for standardized.

Table C2. Lee bounds estimations (NKI)

Lower / Upper Lee bounds estimates

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

Bootstrapped (5000 reps) standard errors in parentheses.
*** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.1
(std) stands for standardized.

Table C3. Forecasting day (NKI)

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, ** p-value < 0.05, * p-value < 0.1

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

Dependent variable	Standardized RFE			
Task	RW (1)	SPmonthly (2)	SPdaily (3)	NKI♠ (4)
CFA Dummy (std)	0.009	-0.018**	0.026**	-0.116**
	(0.036)	(0.008)	(0.013)	(0.050)
Constant	0.231***	-0.025	0.065	-0.020
	(0.057)	(0.059)	(0.055)	(0.080)
R^2	0.034	0.164	0.445	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, ** p-value < 0.05, * p-value < 0.1
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 C2 in Appendix C to correct for selection issues.
(std) stands for standardized.

Table C5. Relative Forecast Error
(SPmonthly treatments, not controlling for CRT)

Dependent variable	Standardized RFE	
Task	SPmonthly (Incongruent) (1)	SPmonthly (Congruent) (2)
CFA Dummy (std)	-0.029**	0.00001
	(0.012)	(0.011)
Constant	-0.114	0.003
	(0.077)	(0.084)
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, ** p-value < 0.05, * p-value < 0.1
(std) stands for standardized.

Table C6. Relative Forecast Error
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 (std)	0.025	-0.021**	0.015	-0.117**
	(0.046)	(0.008)	(0.012)	(0.050)
Constant	0.227***	-0.089*	-0.002	-0.018
	(0.066)	(0.054)	(0.051)	(0.091)
R^2	0.023	0.162	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, ** p-value < 0.05, * p-value < 0.1
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 C2 in Appendix C to correct for selection issues.
(std) stands for standardized.

Table C7. Relative Forecast Error (controlling for CRT)

Dependent variable	Standardized RFE			
Task	RW	SPmonthly	SPdaily	NKI♠
	(1)	(2)	(3)	(4)
CFA Dummy (std)	0.017	-0.022**	0.015	-0.119**
	(0.046)	(0.008)	(0.012)	(0.050)
CRT score (std)	-0.091*	-0.013	-0.002	0.014
	(0.053)	(0.010)	(0.015)	(0.035)
Constant	0.246***	-0.086	-0.002	-0.018
	(0.066)	(0.053)	(0.051)	(0.091)
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, ** *p*-value < 0.05, * *p*-value < 0.1

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 C2 in Appendix C to correct for selection issues.

(std) stands for standardized.

Table C8. Relative Forecast Error
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)
CFA Dummy (std)	0.090*	-0.017*	0.021	-0.044
	(0.050)	(0.009)	(0.013)	(0.051)
Financial literacy (std)[1]	-0.133***	-0.009	-0.013	-0.014
	(0.047)	(0.012)	(0.016)	(0.098)
Constant	0.264***	-0.087	-0.001	0.010
	(0.067)	(0.053)	(0.051)	(0.099)
R^2	0.035	0.161	0.483	0.059
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, ** *p*-value < 0.05, * *p*-value < 0.1
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 C2 in Appendix C to correct for selection issues.
(std) stands for standardized.

[1] 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
(SPmonthly treatments, controlling for financial literacy and CRT)

Dependent variable	Standardized RFE					
Task		SPmonthly (Incongruent)			SPmonthly (Congruent)	
	(1)	(2)	(3)	(4)	(5)	(6)
CFA Dummy (std)	-0.025**	-0.026**	-0.019	-0.016**	-0.016**	-0.015**
	(0.012)	(0.013)	(0.013)	(0.007)	(0.007)	(0.008)
CRT score (std)	-	-0.011	-	-	-0.021*	-
		(0.015)			(0.012)	
Financial literacy (std)[a]	-	-	-0.013	-	-	-0.002
			(0.018)			(0.010)
Constant	-0.158*	-0.156*	-0.156*	-0.072***	-0.068***	-0.072***
	(0.081)	(0.080)	(0.081)	(0.020)	(0.020)	(0.020)
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, ** p-value < 0.05, * p-value < 0.1
(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.