



# Interaction between price and expectations in the jar-guessing experimental market

Toshiaki Akinaga<sup>1</sup> · Takanori Kudo<sup>2</sup> · Kenju Akai<sup>3</sup>

Received: 4 October 2019 / Accepted: 13 October 2022 / Published online: 25 October 2022  
© The Author(s) 2022

## Abstract

This study evaluates the interdependence between pricing and expectations. We investigated not only the ways in which traders' thoughts determined asset prices, but also the feedback process from prices to expectations. In our laboratory market, subjects were asked to estimate the number of balls in a jar and trade an asset whose value was equal to that number. Our asset market, where transactions were eventually settled at the asset value, was like futures markets. The subjects alternately repeated the process of guessing and transaction. A double-auction was used to design our market. Our findings indicated a downward bias in the subjects' estimates, which led to lower transaction prices, since the price converged to the equilibrium price that was determined by the median of estimates. The subjects' experience in our laboratory markets had no systematic effect on the accuracy of estimates, but made them less heterogenous. Our subjects were apt to revise their estimates with reference to prices in a market. We examined the estimation revision process of the subjects using the partial adjustment model.

**Keywords** Experimental asset market · SSW · LtFE · Mispricing · Expectation formation · Wisdom of crowds

---

This work was supported by JSPS KAKENHI Grant Number JP26380409. Tokoha University also supported our study by the Special Research Aid and the Presidential Research Grant. We reported this study at WEHIA 2019 in London under the title “Interaction Between Price and Expectation in Experimental Asset Markets,” a thorough revision of “Reflexivity and Mispricing in Experimental Markets,” presented at ISCEF 2018 in Paris.

---

✉ Toshiaki Akinaga  
akinaga@sz.tokoha-u.ac.jp

<sup>1</sup> Faculty of Business Administration, Tokoha University, 6-1 Yayoi-cho Suruga-ku, Shizuoka City, Shizuoka, Japan

<sup>2</sup> Faculty of Science and Engineering, Setsunan University, 17-8 Ikedanaka-machi, Neyagawa City, Osaka, Japan

<sup>3</sup> Center for Community-based Healthcare Research and Education, Shimane University, 89-1 Enya-cho, Izumo City, Shimane, Japan

**JEL Classification** C92 · D84 · G12 · G41

## 1 Introduction

In the market, traders guess the value of an asset and take a behavior based on their expectations about the asset value. The market disseminates information in the form of, for instance, prices that reflect traders' expectations. Information, such as prices, affects traders' expectations. As George Soros (1987, p. 2) says, cognition of participants and their participation in the market interfere with each other. Soros, a legendary investor, calls such interference *reflexivity*.

Soros (1987, p. 16) complains that economic theories strangely neglect *reflexivity*. However, modern economic studies have attempted to unveil these reflexive aspects by introducing interactive agents explicitly in their models. Learning-to-forecast experiments (LtFEs) were introduced by Marimon et al. (1993) and spawned a series of works: Heemeijer et al. (2009), Bao et al. (2012, 2017), and Colasante et al. (2019) investigated interaction between prices and expectations in the laboratory. In these experiments, subjects forecast a price and their forecasts were aggregated to determine the price; then, they forecast again considering the realized prices. These studies focused on expectation formation rather than price formation, requiring subjects to only forecast prices. LtFEs, based on price adjustment rules, relied on a computer to calculate prices from subjects' forecasts.

In contrast, experimental economics has a long-standing tradition of studying price formation. Double-auction (DA) experiments pioneered by Smith (1962) reveal that simple laboratory markets can discover the price that is efficient by a certain criterion. In these classic experiments, values of tradable objects are given by an experimenter as experimental parameters; therefore, there is no need for subjects to form an expectation about the value of objects. Further, Smith et al. (1988) and Haruvy et al. (2007) investigated expectation formation in DA markets.<sup>1</sup> In these studies, assets brought about stochastic dividends. The probability distribution of dividends, given by an experimenter, determined the asset value.

The present study also investigates expectation formation in DA markets. However, the value of our asset is deterministic and set as an experimental parameter. We used a glass jar filled with small balls to define the asset value. We asked our subjects to guess the number of balls in the jar and to trade an asset whose value is equal to that number. Treynor (1987) conducted a similar jar-guessing experiment in his classroom. The mean estimate of the number of beans in a jar, as estimated by his students, was close to the true value, but when they were cautioned, the mean of their estimates became worse. Treynor (1987) conjectured that his warnings about a jar caused a shared error to creep into individual estimates. This study suggests that the heterogeneity among individual expectations is essential for the accuracy of the average expectation, which determines the price in the market.

Our study aimed to investigate whether the effect of the market on expectations, if any, deteriorates or improves the accuracy of expectations. We also wanted to exam-

<sup>1</sup> Haruvy et al. (2007) used call market rules rather than continuous DA.

ine how subjects' expectations about the asset value determine its price in the market. Our experiment was conducted under the environment of the constant fundamental value, which Noussair et al. (2001) reported as a factor making bubbles smaller and less frequent in experimental markets, although it was not sufficient to eliminate the possibility of a bubble. Such a simple environment would provide a benchmark for future studies. We found that prices converged to the median expectation of the asset value. This convergence seemed to be a necessary consequence of the subjects' behavioral tendency in trading, wherein the subjects with higher expectations tended to buy and those with lower expectations were apt to sell. Our subjects tended to revise their expectations, that is, their estimates of the asset value, in accordance with the market prices. Despite this, our experimental markets showed no definitive effect on the accuracy of the subjects' estimations. This is because their revision process was consistent with the partial adjustment formula that targets the median estimate without any significant disturbances.

We explain the design of our experiment in Sect. 2, and analyze experimental outcomes in Sect. 3, where an analysis on trading behavior is in Sect. 3.2.2. To examine the process of individual subjects to revise their expectations, we use the partial adjustment model in Sect. 4, where Sect. 4.2 presents our analysis of the disturbance in this model. The disturbance has a potential to improve or deteriorate the estimation of subjects as a cohort. Section 5 discusses the notion of the wisdom of crowds, the comparison to previous studies, and the directions for future research. We provide concluding remarks in Sect. 6.

## 2 Experimental design

In our experiment, each subject had two roles: a forecaster and a trader. A forecaster estimates the number of balls in the jar and submit an estimate as a prediction of the value of an experimental asset. A trader transacts to buy or sell an asset in the DA market. We paid our subjects cash rewards according to their performance.

### 2.1 Asset

Our experimental asset—neither dividends nor coupon payments—had a deterministic fundamental value. We called this asset a “certificate”<sup>2</sup> to allow the subjects, many of whom were unfamiliar with finance, to easily understand the experimental procedures. A jar filled with bingo balls defined the value of the asset. After the completion of the period of the experimental market, transactions were settled at the value, i.e., the number of balls in the jar. Therefore, the profits from trading were calculated as follows:

$$\text{Seller's Profit} = \text{Price} - \text{Value} \quad (1)$$

$$\text{Buyer's Profit} = \text{Value} - \text{Price} \quad (2)$$

<sup>2</sup> We used the Japanese word *shouken* for asset in our instructions. *Shouken* is a certificate to claim something. Equities, bonds, and checks are different types of *shouken*.

Acquired profits were converted into cash payments at the end of the experiment.

The transaction of the asset in our laboratory was similar to the futures contract in real markets, though we called the asset “certificate” in our instructions. In real world, traders often close their positions before the expiry by taking opposite positions. Such closeouts through reverse contracts were impossible in our market because we allowed our subjects to trade only once in a period. Therefore, all transactions, calculated as per equations (1) and (2), were settled in cash.

The life of the asset was one experimental period without the possibility of resale or buyback. Therefore, the only way to gain a positive profit was either to buy at a price lower or sell at a price higher than the real value of the asset. Expectations about price movements were almost useless, since there was no chance for capital gain. This allowed our subjects to concentrate on estimating the fundamental value of an asset without bothering themselves about the behavior of other subjects.<sup>3</sup> Experimental markets designed in this study provided minimal to no opportunity for speculation and thus could be used as a benchmark to assess the effect of speculation in speculative markets.

## 2.2 Jar-guessing

We used a 1.7-l transparent glass jar filled with balls to define the value of the asset. The experiment used Japanese style bingo balls of nine colors, each 12-millimeters in diameter, without any numbers on their surface.

Before the experimental market was opened, we showed our subjects a jar and asked them to guess the number of balls in the jar. 24 subjects participated in a session. If they crowded around a jar, it would be difficult for some of them to observe a jar. Therefore, we presented two jars with the same number of balls during the observation period. The subjects could observe either of the two jars revealed by the experimenter. After observation, they submitted their estimates as their predictions about the fundamental value through the computer provided in their cubicles.<sup>4</sup> We instructed all subjects to “submit your prediction within 30 s in general” and mentioned that “the remaining time would be displayed at the upper-right corner of your screen.” However, they could submit their estimates as predictions even after the required 30 s were over. We paid monetary rewards to the subjects based on the precision of their predictions.

Our experiment had no strict controls or restrictions about how the subjects observed the jar to make their estimates. When the experimenter announced that he would present jars, the subjects could leave their cubicles to watch a jar and then return to their seats to enter their estimate of the number of balls in the jar. They could return to their cubicle any time during the observation period. If some subjects continued their observation more than 2 min, the experimenter would gently notify the subjects to return to their seats. For example, he announced to all the subjects that “If you make a guess, you can go back to your seat and click a ready-to-start button.” We covered

<sup>3</sup> It was useful, even in our market, to speculate about other traders’ expectations, in order to determine whether to accept an offered price or wait for a more favorable one. However, such speculations were unnecessary for avoiding a loss in our market.

<sup>4</sup> They could enter an integer between 0 and 999 as their prediction.

the jars after all subjects returned to their seats to avoid providing an unnecessary advantage of unlimited observation time for subjects who were incidentally seated near the jars.

We used this method to alleviate the inequality of observation among subjects and, at the same time, to also avoid an artificial restriction that could possibly hinder natural guessing.

### 2.3 Asset trading

In a session, the subjects experienced nine periods of asset trading but were unaware that the ninth was the last.<sup>5</sup> Each subject could bid or ask freely during a period until they executed a contract. Every subject could choose to be either a buyer or a seller in each period. A subject could trade up to one asset in each period. Once they had bought or sold an asset, they had to wait until the market closed, since reverse contracts of resale or re-purchase were impossible.

Assets were traded using Smith's (1962) DA market system, where buyers' English auction and sellers' Dutch auction proceeded simultaneously. The subjects could execute their contracts once they chose a price from the order book<sup>6</sup> and confirmed to execute the transaction. Our markets were computerized and used Fischbacher's (2007) z-Tree software package.<sup>7</sup> The subjects observed how a market operated on their monitors<sup>8</sup> and submitted their asks and bids using the mouse and keyboard.<sup>9</sup>

Each trading period, separated from a period of jar observation and estimation, lasted for 5 min. The remaining time would be displayed at the upper-right corner of the monitor screen. The period ended before this time limit if no further transactions remained in the market.

### 2.4 Subject

For each session, as participants, we recruited 24 college students of Kyoto Sangyo University. A total of 48 students participated in our two sessions. All of them were

---

<sup>5</sup> We told our subjects that they would alternate between the prediction period and trading period, and a jar that defined an asset value would be shown every three periods of prediction and trading. While the subjects actually observed a jar three times in a session, we only instructed them to observe the same jar *several times*. They cannot estimate the last period from a passage of time, although they were informed about the time schedule at the time of recruitment, because we announced a 3-h schedule for an 80-min session. For an 80-min experiment, a 2-h schedule is generally announced to secure sufficient time for calculating and preparing cash rewards; however, for this experiment, we asked participants to spare 3 h instead of 2 because we had planned a different experiment at the time of recruitment.

<sup>6</sup> The best bid and ask were chosen by default.

<sup>7</sup> Our program for z-Tree is available upon reasonable request.

<sup>8</sup> Each computer screen displayed the order book and transaction prices in sequence during a period. A screen of each subject also displayed his/her own prediction submitted at the last period. His/her own contracted price at a period would also be shown on the screen after his/her transaction.

<sup>9</sup> The subjects could enter a price of an integer between 0 and 999. To submit a new order, they were required to improve the best quote: that is, either raise the bid or lower the ask. They could withdraw their own pending order through a new order or a cancel.

undergraduate. Our recruitment excluded graduate students majoring in economics<sup>10</sup> and foreign students.<sup>11</sup> We had no control over other attributes; hence, our sample contained subjects with varying age, gender, and academic majors.<sup>12</sup> We excluded those who had previously participated in our earlier experiments. Nobody participated in both the sessions. Thus, all subjects were novices to the experimental tasks of this study.

## 2.5 Procedures

Our experiment was conducted in a computerized laboratory. The subjects were seated in a cubicle with a monitor screen, a mouse, and a keyboard. Written instructions, which were also read aloud by an experimenter at the beginning of a session, were distributed to each subject.<sup>13</sup>

After the experimenter read aloud instructions to the subjects, a jar was revealed and presented to the subjects; they were then asked to predict the number of balls in the jar and submit their estimates as predictions. There was no dry run. After submissions were made, the market was opened for trading assets. During a 5-min trading period, each subject could buy or sell an asset. When the trading period ended, they were required to submit a prediction of the number of balls in the jar again, and then the second trading period started. After the third prediction and subsequent trading period, we showed the same jar again without the cover, and then the subjects alternated their prediction and trading thrice. After the sixth prediction and trading, we showed the same jar once more, followed by three predictions and three markets.

In sum, our subjects alternated between jar estimates and asset trading nine times, and they saw the same jar three times. All subjects knew that the predictions and trading would occur three times after they saw the jar, but were unaware about how many times they would be able to observe the jar. We revealed the number of balls in the jar only after the experimental session. Profits obtained through predictions and trading were also announced at the end of the session.

We omitted a dry-run period to exclude prior interaction as far as possible. Our aim was to study the holistic process of inter-participant interaction. However, the trial market entails some interaction among subjects, so we decided to omit it. We also skipped a trial period of jar-guessing, even though we could design a non-interactive trial for it. This is because we wanted to observe how the market affects naive subjects who had no prior training of jar-guessing.

To reduce the noise caused by erroneous tradings, we designed a simple market and carefully instructed our subjects on how to trade in this market and showed them sample screenshots of a trading monitor. In each session, about 20 min was devoted to instruct

---

<sup>10</sup> Friedman and Sunder (1994, Section 4.1.1) recommend excluding doctoral students from economics departments or business schools “because they often respond more to their understanding of possibly relevant theory than to the direct incentives of your laboratory economy.” Taking a more conservative approach, we excluded all graduate students of economics.

<sup>11</sup> We excluded foreign students because of our concern about language and communication.

<sup>12</sup> Table 28 in Appendix F summarizes the school year, gender, and academic major of our subjects.

<sup>13</sup> The instructions for this experiment, in Japanese, are available upon reasonable request.

participants about the procedure and computer-interface, despite which, we anticipated some errors and hesitation regarding decision-making among participants. However, a participant without complete understanding about the experimental market was not necessarily a nuisance in our experiment. Indeed, some participants may disturb a market, but we were also interested in assessing whether the market would absorb or magnify such a disturbance.

## 2.6 Rewards

We paid rewards to participants in cash (JPY), based on their performance. Profits and losses were accumulated and paid in cash at the end of an experimental session.

The reward for a completely accurate prediction was 200 yen, while 1 yen was deducted for every unit error. The minimum reward for a prediction was set at zero; hence, a prediction with an error of 200 or more units yielded no payment.

The profit from asset trading, calculated by formulae (1) and (2), could be negative (i.e., a loss). We endowed each subject with a show-up fee of 750 yen and a trading fund of 2000 yen. If a subject suffered a loss in a market, this initial payment would be reduced.<sup>14</sup> A subject who made no transactions during a market period made neither profit nor loss during that period.

## 2.7 Sessions

We conducted two sessions at the laboratory of Kyoto Sangyo University in Kyoto, Japan. Session 1 was held on 19th February, 2019, and Session 2 on 5th March, 2019. Each session lasted 78 min,<sup>15</sup> including 18 min of instruction, and comprised 24 participants each. The number of balls in the jar, the fundamental value of the asset, was 384. The total payment of rewards, in JPY, was 88 682 and 80 593 for Sessions 1 and 2, respectively.

# 3 Experimental results

## 3.1 Prediction of asset value

### 3.1.1 Downward bias and diminishing diversity

One of our initial findings was that the subjects' predictions regarding the fundamental value—the number of balls in a jar—had a downward bias. Figure 1 shows the mean and median of the 48 predictions submitted in each period of the two sessions. They consistently remained below the true value, that is, 384. Figure 2 is a histogram of

<sup>14</sup> The initial 2750 yen was enough to avoid the bankruptcy problem since there were some earnings from predictions.

<sup>15</sup> A session ended when the ninth market closed, and each achievement was revealed on each respective screen. After a session, the subjects waited in the laboratory for about half an hour because time to prepare cash rewards was necessary.

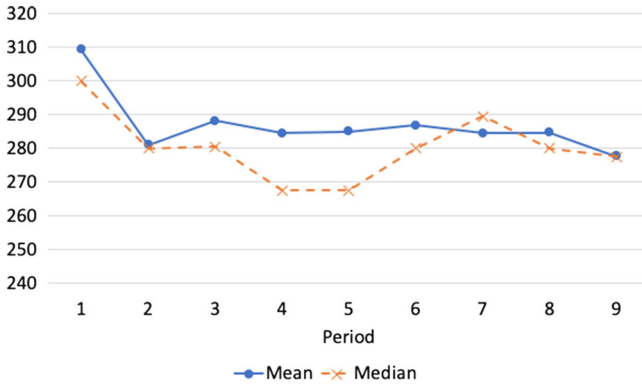


Fig. 1 Mean and median of predictions in the two sessions

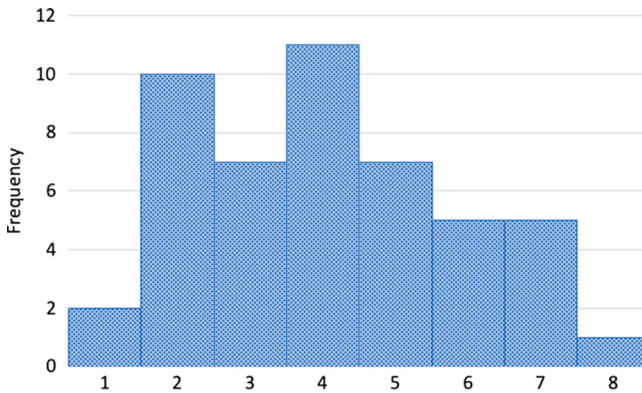


Fig. 2 Prediction distribution at 1st period in the two sessions

predictions made during the first period of the two sessions.<sup>16</sup> The fundamental value 384 lies at the midpoint of the fifth class [346, 422). The mode is immediately to the left of this class. This initial distribution is skewed right with the peak veering to the left; the skewness is 0.208. Our subjects alternated between jar-guessing and asset trading, and their trading experience in the laboratory market seemed to affect their predictions. The frequency table of Appendix B represents how the prediction distribution changed over nine periods in our experiment.

Figures 3 and 4 represent the transition of the cumulative distribution of predictions in Sessions 1 and 2, respectively. Each figure contains three distribution lines for the first, mid, and last periods, that is, the first, fifth, and ninth periods. The lines became gradually step in each session, implying diminishing diversity of predictions. The declining tendency of the standard deviation and the interquartile range in Fig. 5 also indicated that the market made the subjects' predictions less heterogeneous. A decrease

<sup>16</sup> The class width is 76, which is about 20% of the fundamental value. Each class was closed at the lower bound and opened at the upper bound. The first and the eighth class, [42, 118) and [574, 650), contained the minimum and maximum prediction, respectively. The number of observations was 48 as 24 subjects submitted their predictions once during a period in each session.



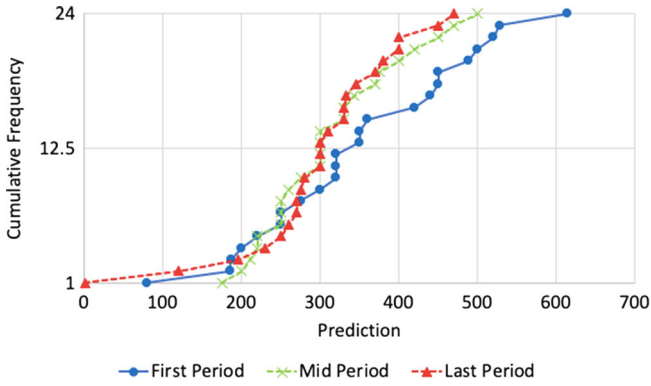


Fig. 3 Cumulative distribution of predictions in Session 1

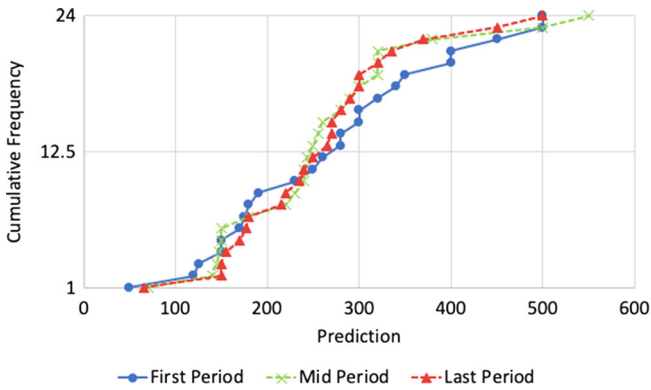


Fig. 4 Cumulative distribution of predictions in Session 2

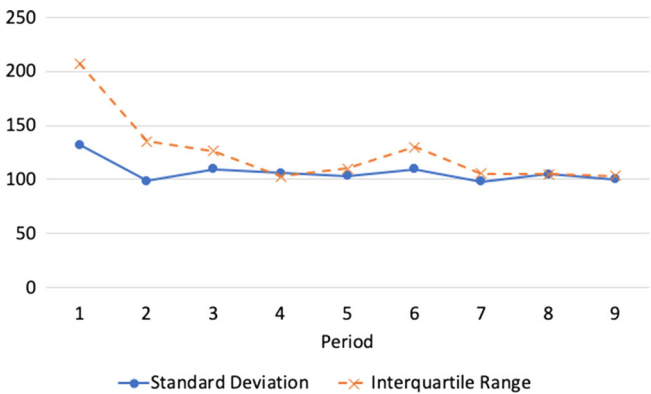
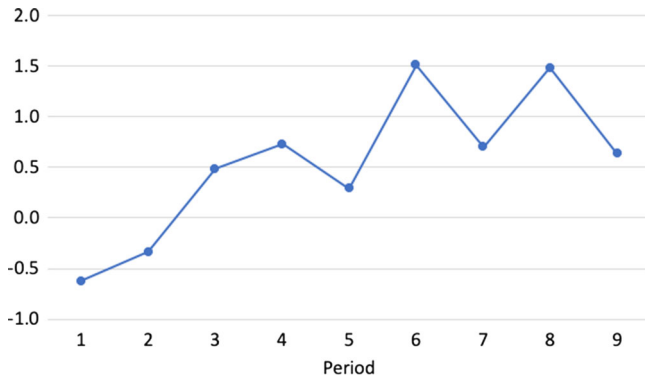


Fig. 5 Standard deviation and interquartile range of predictions in the two sessions



**Fig. 6** Kurtosis of predictions in the two sessions

was clearer in the interquartile range than in the standard deviation. This suggests that some outliers persisted in their own predictions. The increasing kurtosis in Fig. 6 was conformable to the existence of adamant outliers.

Although the divergence in the predictions decreased due to the market experience, it seemed to have had no systematic effect on the precision of the predictions. The mean and median of the predictions changed erratically, as shown in Fig. 1. *T*-statistics of the mean prediction error were always significantly negative. The maximum *T* value, that is, the closest value to zero, was  $-3.913$  at  $t = 1$  with a single-sided significance probability of  $1.461 \times 10^{-4}$ .

The mean and median in Fig. 1 diverged widely during the middle periods.<sup>17</sup> This implies that the prediction distribution was more asymmetric during these periods. The skewness increased in the middle periods, as seen in Table 17 of Appendix A and Fig. 15 of Appendix D. The jar observation may have caused predictions of the fundamental value to be distributed asymmetrically. We consider this possibility in Appendix D.

A table of statistics regarding submitted predictions is presented in Appendix A. Every figure in this subsection except for the histogram, which depicts frequencies in Table 18 in Appendix B, represents the statistical values in Table 17. This table in Appendix A also presents all values of the skewness and *T*-statistics for every period.

### 3.1.2 Time for guessing

Table 1 presents the observation time taken to study the jar. The time for the first observation began when the experimenter started the z-Tree computer program and displayed the jars after he finished reading the instructions. The time for the second and third observation started when the third and sixth period of the market closed,

<sup>17</sup> An anonymous reviewer pointed this out.

**Table 1** Time for jar observation

Obsevation period	Session 1		Session 2	
	Mean	Median	Mean	Median
1st	175.25	172	158.92	158
2nd	175.58	189	136.25	137
3rd	88.42	81.5	124.71	146

The time is in seconds  
24 observations for each period in each session

**Table 2** Time to submit prediction

Period	Session 1		Session 2	
	Mean	Median	Mean	Median
1	16.92	15.5	21.38	21
2	16.46	16	16.83	15
3	9.83	7	9.67	6.5
4	9.50	8	7.79	7
5	5.46	3.5	7.08	4.5
6	5.83	4	8.00	4
7	5.96	5	7.50	6
8	6.92	5	8.13	5
9	7.25	5	6.92	5

The time is in seconds  
24 observations for each period in each session

respectively.<sup>18</sup> An observation time of each subject ended when a subject clicked the ready-to-start button after he/she returned to the cubicle.<sup>19</sup>

Table 2 presents the time taken by subjects to submit their prediction. The time for the first, fourth, and seventh period began when all subjects clicked the ready-to-start button after they came back to their seats. The time for other periods, i.e., for  $t = 2, 3, 5, 6, 8,$  and  $9,$  started when the latest market closed. A time of each subject ended when a subject submitted his/her prediction.<sup>20</sup>

In both Tables 1 and 2, the time taken tended to decline. Learning or fatigue, associated with experimental tasks, could be a reason for this tendency. Some subject might not understand the experimental procedure sufficiently during earlier periods. As periods proceeded, they would become accustomed to the procedure and could be quicker in their decisions and actions. Such an effect of learning was probably important at the earliest periods because we omitted trial periods in our experiment.

The time for the jar observation decreased drastically at the third observation in Session 1, as seen in Table 1. It is possible that the subjects in Session 1 might be quick

<sup>18</sup> The experimenter displayed the jars and asked the subjects to come and watch them again after the third and sixth market closed.

<sup>19</sup> The experimenter made an announcement asking all the subjects to click the ready-to-start button when it was noticed that some subjects had not clicked the button even after all subjects went back to their seats.

<sup>20</sup> The experimenter made an announcement asking all the subjects to submit a prediction when it was noticed that some subjects had not submitted, even though the built-in standard time of 30-s was over.

**Table 3** Number of nonzero revisions and transactions

Period	Session 1		Session 2	
	Nonzero revision	Transaction	Nonzero revision	Transaction
1	19	11	10	10
2	10	11	14	11
3	19	11	19	11
4	13	11	11	11
5	15	10	14	9
6	17	11	20	10
7	13	11	15	9
8	13	11	12	11
9	NA	11	NA	10

24 is the maximum for the nonzero revision at each period in each session

12 is the maximum for the transaction at each period in each session

learners. This characteristic could cause the outcomes of Session 1 to differ from those of Session 2. In addition, the subjects in Session 1 took longer at the first and second observations than did the participants in Session 2. This could make the predictions in Session 1 more accurate than those in Session 2 if longer observations would entail more precise predictions. We discuss these topics in Appendices E and F,<sup>21</sup> although we found no clear correlation between the observation time and prediction accuracy.

During later periods, subjects might have gotten tired and avoided a careful consideration. In fact, one subject in Session 1 submitted the value 1's as predictions for the jar at the eighth and the ninth period. This could be due to either disinterest or fatigue. However, this was merely an exceptional case. We think that the fatigue effect was not so important in our experiment because the activity of the subjects showed no clear tendency to decline (Table 3). Data in Table 3 indicate how many subjects did not vacuously revised their predictions, reporting the number of subjects who did not submit the same prediction as the preceding latest period. The table also contains the transaction volume at each market period. We can confidently state that many of our subjects were not very tired and actively continued revising their estimates or trading an asset even during the later periods.

## 3.2 Transaction prices

### 3.2.1 Price convergence

Prices in our experimental markets stayed below the fundamental value of the experimental asset, reflecting the downward bias of the subjects' expectations about the asset value. Figures 7 and 8 show transaction prices of each session, where the horizontal dashed line represents the asset value defined by the jar. Each figure contains nine segmented lines, corresponding to nine periods of the market. Prices appeared

<sup>21</sup> We are grateful to an anonymous reviewer for suggesting these topics.

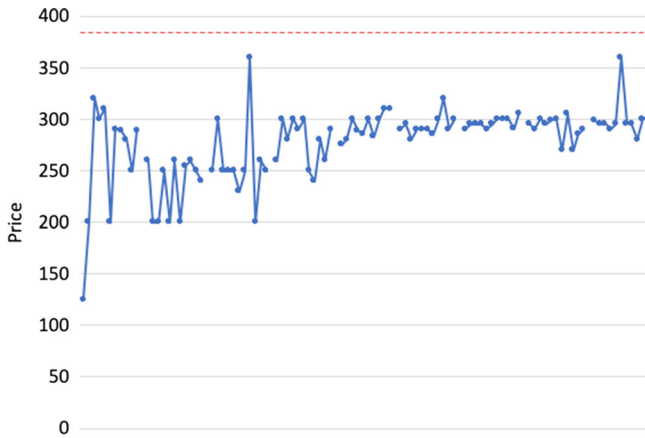


Fig. 7 Price in Session 1

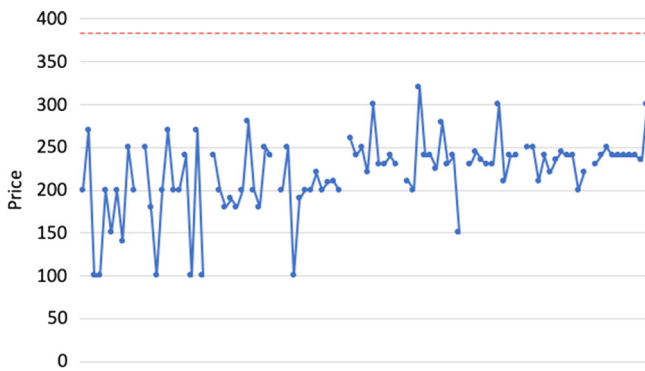


Fig. 8 Price in Session 2

**Table 4** Standard deviation of price

Period	1	2	3	4	5	6	7	8	9
Session 1	60.17	27.64	40.85	21.49	12.46	10.51	4.80	11.72	20.51
Session 2	57.63	66.00	34.09	36.25	24.04	45.23	24.62	16.77	19.78

to fluctuate less sharply as the market periods proceeded. In fact, Table 4 indicates a declining tendency of the standard deviation of the price. The price seemed to converge to a certain level, though it never converged to the fundamental value.

The price of our laboratory markets converged to the equilibrium price, but not to the fundamental value. We defined the equilibrium price at  $t$  by the median of predictions submitted directly before the  $t$ -th market period. The median prediction divided the subjects into two equal-sized groups. One group consists of subjects who predicted relatively larger values. They would buy an asset at the price of the median prediction because the asset was cheaper than their own valuation. In contrast, this

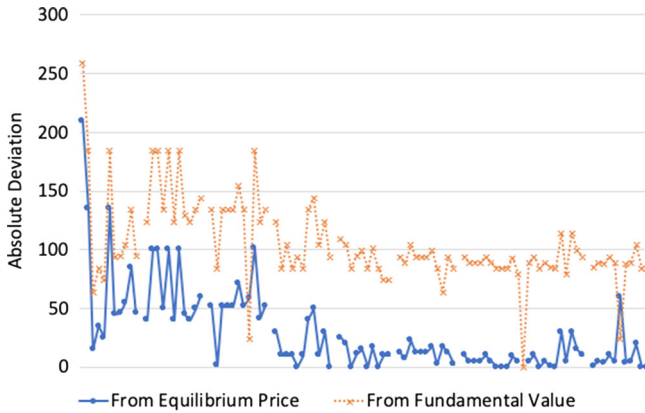


Fig. 9 Absolute deviation of price in Session 1

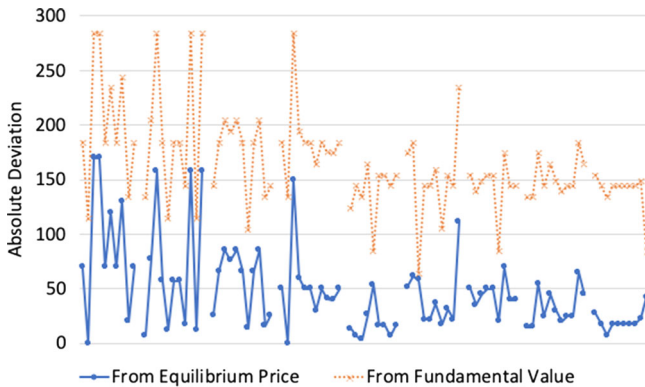


Fig. 10 Absolute deviation of price in Session 2

price was expensive for subjects in the other group who predicted smaller values, and consequently they would sell. Therefore, the demand and supply would balance at the price that was equal to the median prediction, since every subject could trade only one asset in each period. The price of our experiment converged to the equilibrium price defined by the median prediction. Figures 9 and 10 show the absolute deviation of the transaction price from the equilibrium price, which declined to near-zero. Each panel (Figs. 9, 10) also involves the absolute deviation from the fundamental value, which is rather distant from zero. Prices were attracted to the median prediction which was always significantly smaller than the fundamental value 384.<sup>22</sup> Table 5 presents equilibrium prices, that is, the medians of the subjects' asset valuations, at every period in each session.

In traditional DA experiments, the demand and supply are strictly controlled through parameters like values/costs that clearly define at what prices each subject could obtain a positive profit. In our experiment, expectations set by the subjects induced the demand

<sup>22</sup> Prices were often closer to the fundamental value and more convergent to the equilibrium price in Session 1 than in Session 2. We discuss this in Appendix F.

**Table 5** Equilibrium price/median prediction

Period	1	2	3	4	5	6	7	8	9
Session 1	335	300	301.5	290	300	302.5	300	300	300
Session 2	270	257.5	266	250	246.5	261.5	280	265	257.5

24 observations for each period in each session

and supply. We controlled them only indirectly, through the jar with balls. However, prices in our DA markets exhibited a good convergence to the equilibrium price.

### 3.2.2 Auctioning behavior

In our experimental markets, the subjects were able to trade only one asset in each period. Therefore, the demand and supply would be balanced, if half of the subjects wanted to buy and the other half wanted to sell. The median of submitted predictions divides subjects into two equal-sized groups. The demand and supply would equilibrate at the price of the median prediction, so long as subjects who submitted lower predictions than the price in the market wanted to sell and subjects of higher predictions wanted to buy.

In our experiment, any subjects could submit either a bid or an ask at their will during each period. They could switch their orders from buying to selling or vice versa at any time so long as their orders were not executed. If just a small fraction of the subjects wanted to sell and an overwhelming majority persisted in buying, the price in the market could be considerably higher than the median prediction. Now, we consider an extreme case where only one subject who submitted the second largest prediction wanted to sell. All other subjects specialized in buying and would never sell even at significantly higher prices than their predictions. In this case, only one asset would be transacted at the price between the largest and the second largest predictions.

Such an imbalance did not occur in our experiment. Let us define a proper buyer and a proper seller as follows. A subject who only bids during a market period is a proper buyer for that period. A proper seller only asks in a period. Some subjects submit both buy and sell orders or no orders within a certain period.<sup>23</sup> They are neither a proper buyer nor a proper seller. Table 6 shows the number of proper buyers and proper sellers at each period. Every period had a balanced number of each type.

Proper buyers usually predicted higher values, and proper sellers tended to estimate lower values. Table 7 summarizes the average rank of prediction according to the subject type. The smallest prediction is ranked 1, and the largest is ranked 24. In case of a tie, each prediction is allocated a mean rank. For example, if the smallest predictions are submitted by two subjects and other 22 predictions are larger, each of the smallest predictions is ranked 1.5. The average rank of predictions submitted by proper buyers was larger than that of predictions submitted by proper sellers at every period in both sessions.

<sup>23</sup> We regard here the acceptance of a price on the order book as a type of bid/ask. Accepting an offered price to buy an asset is a bid. Acceptance of a bid price in order to sell is an ask.

**Table 6** Number of proper buyers and proper sellers

Period	Session 1		Session 2	
	Buyer	Seller	Buyer	Seller
1	11	11	10	11
2	11	8	13	9
3	11	9	12	11
4	9	10	11	11
5	9	9	9	8
6	11	10	11	11
7	10	11	10	8
8	9	12	11	11
9	11	11	11	10

**Table 7** Average rank of predictions by proper buyers and proper sellers

Period	Session 1		Session 2	
	Buyer	Seller	Buyer	Seller
1	14.5	11.2	16.7	8.6
2	16.5	8.7	15.2	8.4
3	15.3	8.4	16.4	7.4
4	18.1	5.7	18.6	6.6
5	18.1	6.2	15.6	6.0
6	18.3	5.9	17.6	6.8
7	18.0	7.3	17.1	4.5
8	17.4	8.0	16.5	8.5
9	16.5	8.1	15.8	6.9

Approximately half of the subjects wanted to buy and the rest wanted to sell. Subjects with higher expectations about the asset value were apt to bid, and subjects with lower expectations tended to ask. Hence, it was not surprising that the prices in our experimental markets converged at the median of predictions, although we need more studies to elucidate the exact mechanism of this convergence.

### 3.3 Interaction between prices and predictions

The divergence in expectations about the asset value decreased in our experiment, as shown in Sect. 3.1.1. Such a clustering of predictions occurred because all the subjects watched the same market, and tended to revise their predictions by referring to prices in the market.<sup>24</sup>

In fact, the revision of the prediction and the gap between the average price and the prediction were negatively correlated. A revision is the difference in predictions of a subject from one period to the next. A gap is the difference between a subject's

<sup>24</sup> We can regard this tendency as a form of anchoring effect (Tversky and Kahneman 1974) as the most recent information received by subjects before revising their predictions was related to prices in the market.



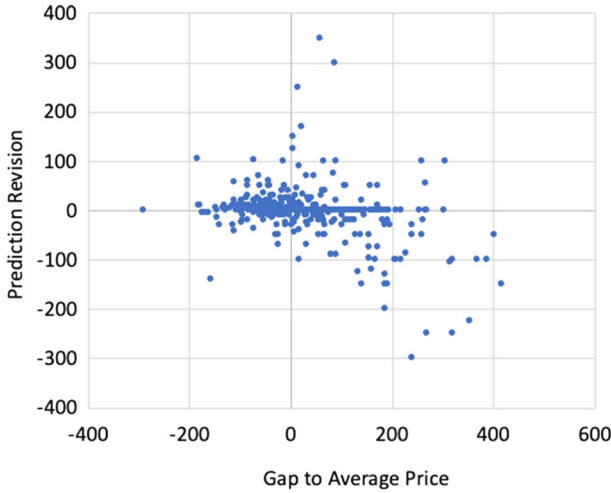


Fig. 11 Gap and revision of predictions in the two sessions

prediction and the average transaction price. A subject submitted a prediction and participated in trading in the following market period and then revised the prediction. This revision appeared to be based on the prices during the trading period directly before a revision. Figure 11 is a scatter diagram of the pooled predictions of the two sessions. The horizontal axis is the gap between the  $t$ -th prediction and the average price of the  $t$ -th market period. The vertical axis is the prediction revision form  $t$  to  $t + 1$ . The correlation was  $-0.337$ .

A regression is useful for analyzing how the subjects revised their predictions, being based on the average price. We regressed the revision on the gap. The regression model is as follows.

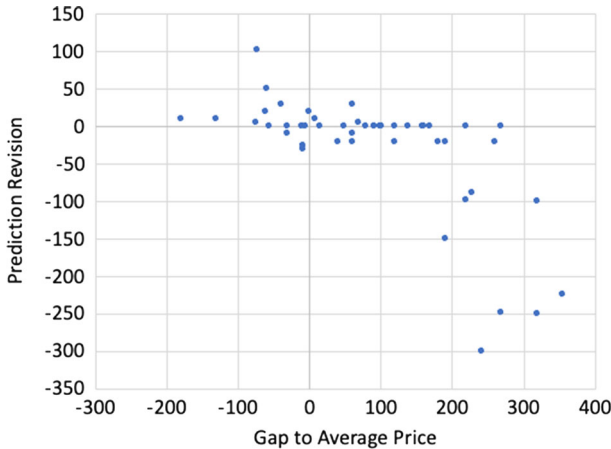
$$V_{i,t+1,k}^e - V_{i,t,k}^e = a + b(V_{i,t,k}^e - \bar{P}_{t,k}) + v_{i,t,k} , \tag{3}$$

where  $V_{i,t,k}^e$  is a  $t$ -th prediction for the asset value submitted by the subject  $i$  in Session  $k$ .  $\bar{P}_{t,k}$  is the average price of the  $t$ -th market of Session  $k$ . If  $n$  transactions occurred at the  $t$ -th market of Session  $k$ ; thus,  $\bar{P}_{t,k} = (\sum_{j=1}^n P_{j,t,k})/n$  where  $P_{j,t,k}$  is the  $j$ -th transaction price of the  $t$ -th market of Session  $k$ . The perturbation is denoted by  $v_{i,t,k}$ . The explained variable is the revision of prediction, i.e., the difference in predictions between adjacent periods for each subject. The explanatory variable, the gap, is the discrepancy between the prediction of a subject before a revision and the average price in the market directly before a subject revised the prediction. As a subject was required to submit nine predictions in a session, each subject revised a prediction eight times. The average price of the ninth market is not used in our regression because no prediction was submitted after the ninth market.

**Table 8** Pooled regression of the two sessions

Intercept			Regression coefficient			$\sqrt{R^2}$
$\hat{a}$	<i>T</i> value	<i>P</i> value	$\hat{b}$	<i>T</i> value	<i>P</i> value	
3.325	1.164	0.245	-0.174*	-6.992	$1.214 \times 10^{-11}$	0.337

Estimates significantly different from 0 at 5% level are denoted by \*  
*P* value indicates a double-sided significance probability  
 The number of observations is 384



**Fig. 12** Gap and revision of predictions in the two sessions of Period 1

Table 8 reports a pooled regression with 384 observations,<sup>25</sup> where  $t = 1, 2, \dots, 8$ ,  $i = 1, 2, \dots, 24$ , and  $k = 1, 2$ . The intercept  $\hat{a}$  was not significantly different from zero. The coefficient  $\hat{b}$  was significantly negative. This means the subjects usually revised their predictions upward if their predictions were below an average price and downward if their predictions were above the average. However, the value of  $\hat{b}$  implies that the ordinary revision was only by a value of 17% of the gap between a prediction and the average price. The value of  $\sqrt{R^2}$ , which is equivalent to the absolute value of correlation, was low, suggesting this revision process was often considerably disturbed.

Table 9 shows a pooled regression with 336 observations, which excludes 48 observations at the 1st period of both sessions. In this regression,  $t = 2, 3, \dots, 8$ . We omitted the trial period in our experimental procedure; hence, our subjects might have been confused at the first period. However, the results from this additional regression were essentially same as the regression presented in Table 8 in that  $\hat{a}$  was not signifi-

<sup>25</sup> The number of observations 384 coincided with the fundamental value. This was just an incident. We fixed the fundamental value parameter according to the number of available balls, on the one hand. On the other hand, the capacity of the laboratory where we did some early experiments, from which the experiment in this paper was derived, restricted the number of subjects to 24. We adjusted the number of periods in order to complete a session within a moderate time.

**Table 9** Pooled regression without observations at 1st period

Intercept			Regression coefficient			$\sqrt{R^2}$
$\hat{a}$	<i>T</i> value	<i>P</i> value	$\hat{b}$	<i>T</i> value	<i>P</i> value	
3.421	1.214	0.226	-0.111*	-4.253	$2.738 \times 10^{-5}$	0.227

Estimates significantly different from 0 at 5% level are denoted by \*  
*P* value indicates a double-sided significance probability  
 The number of observations is 336

cantly different from zero and  $\hat{b}$  was significantly negative.<sup>26</sup> Both  $|\hat{b}|$  and  $\sqrt{R^2}$  were smaller in Table 9 than in Table 8 because some outliers with great revisions were excluded from the additional regression. Figure 12, a scatter diagram at  $t = 1$ , shows observations which were excluded from the regression presented in Table 9.

## 4 Estimation for individual prediction revision

### 4.1 Partial adjustment model

A pooled regression in Sect. 3.3 suggests our subjects revised their predictions by referring to prices in the market. We now examine how an individual subject revised predictions. Each subject’s point of reference probably differed. However, it was inevitably related to prices in a market, and the price in our laboratory markets tended to converge to the median of predictions, as shown in Sect. 3.2.1. Therefore, the median of submitted predictions would be a good approximation to the individual reference in the revision process, even though none of the subjects directly observed the median.

Thus, we start our analysis with the following regression equation for the subject  $i$ ,

$$\Delta V_{i,t}^e = \alpha_i + \beta_i(\mathcal{M}_t - V_{i,t}^e) + e_{i,t} , \tag{4}$$

where  $V_{i,t}^e$  is a  $t$ -th prediction by the subject  $i$ ,  $\Delta V_{i,t}^e = V_{i,t+1}^e - V_{i,t}^e$ , and  $\mathcal{M}_t$  is the median of predictions submitted at the period  $t$ .<sup>27</sup> Every subject submitted nine predictions; then, they revised their predictions eight times at  $t = 1, 2, \dots, 8$ . The degree of freedom was only six; hence, our regression analysis of individual subjects would present just a sketchy description rather than a precise estimation about the subjects’ behavior.

We obtained 48 regression equations from the two sessions as 24 subjects participated in each session. (Tables of estimates are in Appendix C.) Estimates of the regression coefficient,  $\hat{\beta}_i$ ’s, often lay between 0 and 1. Only 13 out of 48 estimates were outside the (0, 1) open interval. Most estimates of the intercept,  $\hat{\alpha}_i$ ’s, were non-

<sup>26</sup> The regression with 288 observations, where both the first period and second period were deleted, yielded similar results. The intercept  $\hat{a} = 1.533$  was not significantly different from zero, and  $\hat{b} = -0.110$  was significantly negative with the double-sided *P* value 0.569 and  $1.559 \times 10^{-5}$ , respectively.  $\sqrt{R^2} = 0.252$ .

<sup>27</sup> We omit the subscript  $k$  to denote the session as we deal with two sessions separately in this section.

significant. At the 5% level, only nine intercepts were significantly different from zero. Values of residuals,  $\hat{e}_{i,t}$ 's, were not very large. (See Sect. 4.2 for details.) Thus, most of our subjects seemed to revise their predictions in accordance to the following partial adjustment process.

$$\Delta V_{i,t}^e = \beta_i (\mathcal{M}_t - V_{i,t}^e), \tag{5}$$

where  $0 < \beta_i < 1$ . This coincides with an adaptive expectation model if  $\mathcal{M}_t = V$ . Therefore, the above partial adjustment model implies that individual subjects would form adaptive expectations, if the median of individual predictions gave perfect foresight about the asset value  $V$ .

Now, we transform (5) as follows, in order to see how predictions could be revised.

$$V_{i,t+1}^e - \mathcal{M}_t = (1 - \beta_i) \{V_{i,t}^e - \mathcal{M}_t\} \tag{6}$$

It is easily seen that  $V_{i,t+1}^e$  is closer to  $\mathcal{M}_t$  than  $V_{i,t}^e$ , and  $V_{i,t+1}^e$  exceeds  $\mathcal{M}_t$  neither from above nor from below  $\mathcal{M}_t$ , since  $1 > 1 - \beta_i > 0$ . Revised predictions get closer to, but never overshoot,  $\mathcal{M}_t$ . Therefore, the revised median  $\mathcal{M}_{t+1}$  stays in the neighborhood of the previous median  $\mathcal{M}_t$ . Now, we introduce a notation about the order of predictions, in order to formalize this argument.  $V_{(n),t}^e$  denotes the  $n$ -th smallest prediction at the period  $t$ . If  $m < n$ , then  $V_{(m),t}^e \leq V_{(n),t}^e$ . This second inequality is an inclusive inequality, since we gave a different placing number for every prediction even when ties exist.  $\mathcal{M}_t$ , the median of predictions at the period  $t$ , is the midpoint of  $V_{(12),t}^e$  and  $V_{(13),t}^e$ , since each experimental session involved 24 subjects. Thus, we have the next proposition.

**Proposition 1** *If all subjects revise their prediction according to the partial adjustment formula (5), then the revised median  $\mathcal{M}_{t+1}$  lies between  $V_{(12),t}^e$  and  $V_{(13),t}^e$ .*

**Proof** The equation (6) that is equivalent to (5) never revise predictions beyond the target  $\mathcal{M}_t$ , since  $1 - \beta_i > 0$ . This implies revisions of  $V_{(1),t}^e, V_{(2),t}^e, \dots, V_{(12),t}^e$  are smaller than or equal to  $\mathcal{M}_t$ , and revisions of  $V_{(13),t}^e, V_{(14),t}^e, \dots, V_{(24),t}^e$  are greater than or equal to  $\mathcal{M}_t$ . Therefore, the lower half predictions at the period  $t + 1$  consist of revisions of  $V_{(1),t}^e, V_{(2),t}^e, \dots, V_{(12),t}^e$ .

Suppose the subject  $i$  submitted the 12th smallest prediction at the period  $t$ , i.e.,  $V_{i,t}^e = V_{(12),t}^e$ .  $V_{i,t+1}^e \leq V_{(12),t+1}^e$ , since  $V_{(12),t+1}^e$  is the largest in lower half predictions at  $t + 1$ . Further,  $V_{i,t}^e \leq V_{i,t+1}^e$  because equation (6) adjusts a prediction closer to the target, as  $1 > 1 - \beta_i > 0$ . Therefore,  $V_{(12),t}^e = V_{i,t}^e \leq V_{i,t+1}^e \leq V_{(12),t+1}^e$ . After all,  $V_{(12),t}^e \leq V_{(12),t+1}^e$ .

$V_{(13),t+1}^e \leq V_{(13),t}^e$  follows in the same manner. Thus, we have the following inequalities.

$$V_{(12),t}^e \leq V_{(12),t+1}^e \leq V_{(13),t+1}^e \leq V_{(13),t}^e$$

This completes a proof, since  $\mathcal{M}_{t+1}$  lies between  $V_{(12),t+1}^e$  and  $V_{(13),t+1}^e$ .

□

Proposition 1 implies that the partial adjustment formula (5) confines the median of revised predictions  $\mathcal{M}_{t+1}$  in a neighborhood of the previous median  $\mathcal{M}_t$ , which is a midpoint of  $V_{(12),t}^e$  and  $V_{(13),t}^e$ . Some deviations from (5) are necessary to revise a median prediction beyond an interval  $[V_{(12),t}^e, V_{(13),t}^e]$ . This means the disturbance to (5) is inevitable not only to perturb appropriate predictions, but also to correct a bias of initial predictions.

Experimental outcomes entailed deviations from (5) more or less, whether temporal or persistent. Even if our subjects followed the partial adjustment formula, their target could deviate from the median prediction, depending on perturbations in transaction prices and variations about their focal points in the market. Now, we introduce  $\epsilon_{i,t}$  to represent a variability in the target, and  $u_{i,t}$ , which denotes the disturbance to the revision process. Then we have the following generalized model.

$$\Delta V_{i,t}^e = \beta_i \{(\mathcal{M}_t + \epsilon_{i,t}) - V_{i,t}^e\} + u_{i,t} \quad (7)$$

$$= \beta_i \{\mathcal{M}_t - V_{i,t}^e\} + \beta_i \epsilon_{i,t} + u_{i,t} \quad (8)$$

Thus, we have the next general model of prediction revision.

$$\Delta V_{i,t}^e = \beta_i (\mathcal{M}_t - V_{i,t}^e) + \eta_{i,t}, \quad (9)$$

where  $\eta_{i,t} = \beta_i \epsilon_{i,t} + u_{i,t}$ .

## 4.2 Analysis of disturbances

The disturbance term  $\eta_{i,t}$  in a general adjustment model (9) is a driving force to correct or exacerbate a bias in initial predictions. The next proposition follows directly from Proposition 1.

**Proposition 2** *If  $\mathcal{M}_{t+1} \notin [V_{(12),t}^e, V_{(13),t}^e]$ , then  $\eta_{i,t} \neq 0$  for some  $i$ , so long as all subjects follow (9) with positive adjustment coefficient  $\beta_i$  that is smaller than one.*

Medians of predictions in our experiment showed only moderate fluctuations and no tendency to increase or decrease (see Fig. 13), suggesting  $\eta_{i,t}$  did not have enough power to mitigate or reinforce a downward bias of initial predictions.

We estimated the disturbance term  $\eta_{i,t}$  in the prediction adjustment model (9) by estimates of the intercept and residuals of regression model (4) as follows.

$$\hat{\eta}_{i,t} = \hat{\alpha}_i + \hat{e}_{i,t} \quad (10)$$

Statistics about  $\hat{\eta}_{i,t}$ 's are summarized in Table 10. The magnitudes of disturbances were not very large. Most  $\hat{\eta}_{i,t}$ 's were in a range of  $0 \pm \mathcal{M}_t/4$  as is shown in Table 11. Table 12 reports Durbin-Watson ratio of regression for every subject. Positive serial correlations were not detected in  $\hat{e}_{i,t}$ 's. Therefore,  $\hat{\eta}_{i,t}$ 's also had no positive serial correlations. This implies that there were not enough accumulated effects of disturbances to drive predictions in one direction.

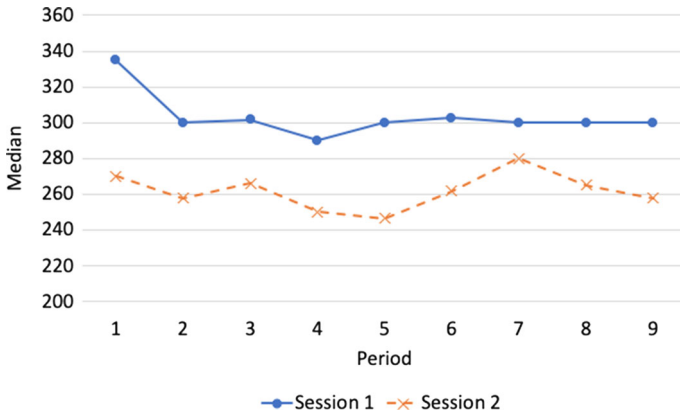


Fig. 13 Medians in each session

Table 10 Statistics of disturbance

<i>t</i>	Session 1				Session 2			
	Avg	SD	Max	Min	Avg	SD	Max	Min
1	4.98	58.18	159.46	-119.14	-2.36	47.85	105.64	-81.11
2	22.19	71.30	219.92	-92.03	9.70	83.54	339.68	-109.60
3	16.37	72.30	159.31	-128.37	8.89	113.93	359.37	-129.95
4	26.01	72.14	189.85	-125.57	13.77	93.97	343.84	-112.74
5	14.09	65.40	177.98	-132.01	22.94	114.78	358.01	-135.66
6	10.19	62.09	157.70	-124.27	14.62	95.45	383.02	-98.60
7	13.49	70.39	161.74	-151.92	6.83	90.23	287.03	-109.35
8	14.94	61.82	163.99	-115.57	0.42	77.19	254.44	-107.84

24 observations for each period in each session

Table 11 Frequency of disturbance in  $0 \pm \alpha \mathcal{M}_t$  Region

<i>t</i>	Session 1			Session 2		
	$\mathcal{M}_t$	$\alpha = 0.1$	$\alpha = 0.25$	$\mathcal{M}_t$	$\alpha = 0.1$	$\alpha = 0.25$
1	335	12	21	270	12	21
2	300	11	18	257.5	14	20
3	301.5	11	19	266	12	17
4	290	12	18	250	13	20
5	300	13	20	246.5	12	17
6	302.5	13	19	261.5	14	18
7	300	13	17	280	16	20
8	300	15	19	265	14	18

$0 \pm \alpha \mathcal{M}_t$  region is a closed interval  $[-\alpha \mathcal{M}_t, \alpha \mathcal{M}_t]$

The maximum frequency is 24 for each period in each session

**Table 12** Durbin–Watson ratio of individual regression

Subject	Session 1		Session 2	
	DW	Ser Corr	DW	Ser Corr
1	2.279	Zero	2.046	Zero
2	2.309	Zero	2.930	
3	2.204	Zero	1.210	
4	2.812		2.273	Zero
5	2.239	Zero	1.971	Zero
6	1.626	Zero	2.170	Zero
7	2.166	Zero	1.699	Zero
8	1.832	Zero	2.934	
9	3.371	Negative	1.567	Zero
10	1.679	Zero	2.587	Zero
11	1.367	Zero	NA	
12	2.515	Zero	1.888	Zero
13	1.876	Zero	2.503	Zero
14	2.789		2.867	
15	2.825		2.127	Zero
16	2.323	Zero	2.856	
17	2.324	Zero	1.550	Zero
18	0.794		2.578	Zero
19	NA		2.249	Zero
20	3.348	Negative	2.108	Zero
21	1.995	Zero	1.218	
22	1.642	Zero	1.699	Zero
23	1.878	Zero	2.236	Zero
24	1.508	Zero	1.843	Zero

$dL = 0.763$ ,  $dU = 1.332$ ,  $4 - dU = 2.668$ ,  $4 - dL = 3.237$

5% significance level

Residuals are zero-constant for Subject 19 and 11 in Sessions 1 and 2, respectively

Table 13 shows the correlation between  $\hat{\eta}_{i,t}$  and the submitted prediction  $V_{i,t}^e$ . They were positively correlated for all  $t$ . This implies that a revision of a larger prediction was often accompanied by a positive  $\hat{\eta}_{i,t}$ , and negative ones tended to occur when smaller predictions were revised. Therefore,  $\hat{\eta}_{i,t}$  was apt to be positive or negative, when  $\mathcal{M}_t - V_{i,t}^e$  was negative or positive, respectively. This means  $\hat{\eta}_{i,t}$  often had an effect that made a new prediction move away from, rather than closer to, the previous median  $\mathcal{M}_t$ . The disturbance term  $\eta_{i,t}$  added inertia, rather than agility, to the prediction revision process in our experiment.

### 4.3 Over-adjustment and overshoot

The disturbance term  $\eta_{i,t}$  did not have enough power to correct or exacerbate a prediction bias in our experiment. Although Proposition 2 insists the power of  $\eta_{i,t}$  is

**Table 13** Correlation between disturbance and prediction

<i>t</i>	Session 1	Session 2
1	0.35	0.89
2	0.39	0.39
3	0.27	0.47
4	0.75	0.87
5	0.74	0.84
6	0.82	0.70
7	0.74	0.84
8	0.65	0.81

24 observations for each period in each session

**Table 14** Non-partial adjustment

Session 1				Session 2			
Subject	$\hat{\beta}_i$	<i>T</i> value	<i>P</i> value	Subject	$\hat{\beta}_i$	<i>T</i> value	<i>P</i> value
<b>2</b>	1.617*	11.752	0.000	<b>3</b>	1.375*	3.321	0.016
3	-0.273	-0.605	0.567	<b>7</b>	1.123*	4.050	0.007
<b>6</b>	1.237*	2.971	0.025	11	0.000	NA	NA
<b>7</b>	1.096*	15.951	0.000	13	-0.798	-1.728	0.135
<b>14</b>	1.030*	2.663	0.037				
<b>18</b>	1.187*	2.822	0.030				
19	0.000	NA	NA				
<b>21</b>	1.148*	3.144	0.020				
22	-0.139	-0.420	0.689				

Estimates significantly different from 0 at 5% level are denoted by \*

*P* value indicates a double-sided significance probability

Over-adjusting subjects are denoted by bold letters

Subjects 19 and 11 in Sessions 1 and 2, respectively, never changed their predictions

essential to release  $\mathcal{M}_{t+1}$  from the neighborhood of  $\mathcal{M}_t$ , it assumes the partial adjustment, i.e.,  $0 < \beta_i < 1$ . If  $\beta_i > 1$ , the prediction revision process of (9) even with zero disturbances could extrude  $\mathcal{M}_{t+1}$  out of  $[V_{(12),t}^e, V_{(13),t}^e]$ . Actually, some subjects followed a non-partial adjustment as is shown in Table 14. Adjustment processes of nine subjects in Session 1 and four subjects in Session 2 were non-partial, i.e., they contradicted the parameter restriction  $0 < \hat{\beta}_i < 1$ . Six and two of them in Sessions 1 and 2, respectively, over-adjusted their predictions, i.e.,  $\hat{\beta}_i > 1$ .

Although the median prediction fluctuated moderately as is shown in Fig. 13, the absolute value of  $\Delta\mathcal{M}_t = \mathcal{M}_{t+1} - \mathcal{M}_t$  was relatively large in Session 1 at  $t = 1$ . (See Fig. 14.) This is partly because Session 1 involved over-adjusting subjects much more than Session 2. In spite of this,  $|\Delta\mathcal{M}_t|$ s were smaller in Session 1 than in Session 2 during every period except  $t = 1$  and 4, in contrast to the largest value at  $t = 1$  in Session 1. Six over-adjusting subjects by themselves were not sufficient to explain



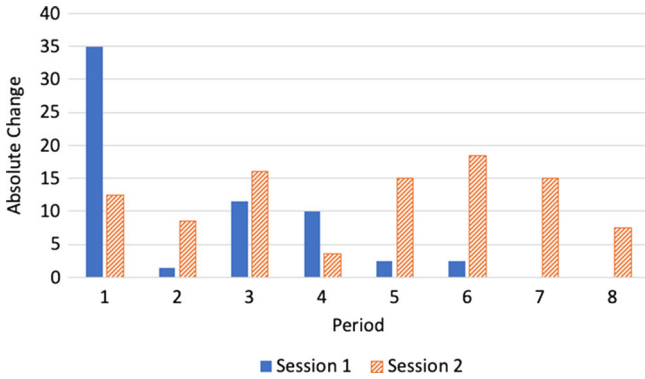


Fig. 14 Absolute change of median in each session

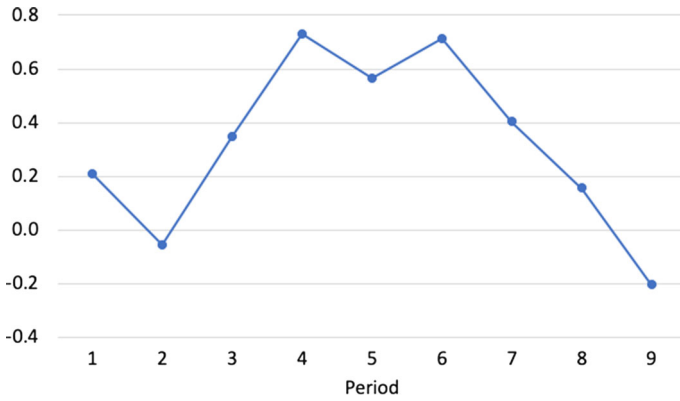


Fig. 15 Skewness of predictions in the two sessions

why  $|\Delta \mathcal{M}_1|$  in Session 1 was the largest. A synergy of the over-adjusting subjects and the effect of disturbances is a key to reveal the reason.

As Table 13 indicates, correlation between the prediction  $V_{i,t}^e$  and the disturbance  $\eta_{i,t}$  was relatively small at  $t = 1$  of Session 1, suggesting weaker inertial effects of the disturbance. Therefore, in this period, reinforcement of adjustment by  $\eta_{i,t}$  was likely. If this reinforcement occurred to over-adjusting subjects, the revision of their predictions would be overshooting. An overshoot is the case when either of the next two conditions holds.

$$0 < \mathcal{M}_t - V_{i,t}^e < \Delta V_{i,t}^e \tag{11}$$

$$0 > \mathcal{M}_t - V_{i,t}^e > \Delta V_{i,t}^e \tag{12}$$

The magnitude and the direction of overshooting were calculated by the next formula.

$$\Delta V_{i,t}^e - (\mathcal{M}_t - V_{i,t}^e) = V_{i,t+1}^e - \mathcal{M}_t \tag{13}$$

**Table 15** Aggregation and frequency of overshoot

<i>t</i>	Session 1			Session 2		
	Aggregated value	Frequency		Aggregated value	Frequency	
		Negative overshoot	Positive overshoot		Negative overshoot	Positive overshoot
1	-210	3	1	-20	1	0
2	233	2	2	342.5	0	1
3	71.5	3	2	-60	4	1
4	20	0	2	0	0	0
5	10	0	1	33.5	0	1
6	-27.5	2	1	65.5	1	2
7	0	0	0	-50	1	0
8	0	0	0	-15	1	0

**Table 16** Frequency distribution of six over-adjusting subjects in Session 1

<i>t</i>	1	2	3	4	5	6	7	8	9
Subjects who predicted larger than $\mathcal{M}_t$	5	2	4	2	2	2	2	4	3
Subjects who predicted just $\mathcal{M}_t$	0	2	0	0	2	0	3	0	1
Subjects who predicted smaller than $\mathcal{M}_t$	1	2	2	4	2	4	1	2	2

Table 15 summarizes the aggregated value of (13) for all overshooting subjects, i.e., for all *i* which satisfies (11) or (12), at each period.<sup>28</sup>

At *t* = 1 in Session 1, three subjects overshoot  $\mathcal{M}_1$  from above, and only one did so from below this target. (See ‘Frequency’ of Table 15.) Further, these overshoots aggregated to -210. (See ‘Aggregated Value’ of Table 15.) The first period of Session 1 showed clearly asymmetric overshooting in both aspects of the magnitude and the frequency. This period, as shown in Table 16, also involved an uneven distribution of six over-adjusting subjects, i.e., five of them predicted above  $\mathcal{M}_1$ , probably just by chance. This transient imbalance, which led to prominently asymmetric overshooting in prediction revision, seems to be a fundamental reason for the largest absolute change of the median prediction at *t* = 1 in Session 1.

## 5 Discussion

### 5.1 Wisdom of crowds

An aggregation of estimates is not uncommonly more precise than individual estimates. Galton (1907), Lorge et al. (1958), and Treynor (1987) showed the precision of

<sup>28</sup> The positive overshoot occurs if the inequalities (11) hold. The negative overshoot is the case in which the condition (12) is satisfied. The overshoot is positive or negative according to whether the value of (13) is positive or negative.

collective estimate of many individual estimates, where each estimate was usually not precise in comparison to the estimation as a cohort, i.e., the mean/median of individual estimates. Surowiecki (2004) cited these studies to demonstrate examples of *the wisdom of crowds*, the notion also known as *the collective intelligence*. The market mechanism is possibly an example of this kind of wisdom, since the market price reflects various insights of many traders. Stock prices in real markets seem to fully reflect available information, and for individual investors to persistently outperform the market index is very difficult (see, e.g., Cowles 1933; Fama 1970; Carhart 1977). Surowiecki (2004) also described the market mechanism as a typical example of *the wisdom of crowds*, while referring to an empirical study of Maloney and Mulherin (2003), who reported that the stock price of the firm responsible for the crash of the space shuttle Challenger had plunged most sharply soon after the crash, long before the cause of the accident was determined.

Contrary to the results of these previous studies, our subjects, as a cohort, underestimated the number of balls in the jar, and prices in our laboratory markets never revealed the true value of the asset. The results could have been influenced by cultural bias as our experiment was conducted in Japan, with only Japanese subjects. According to Rieger et al. (2014), who had surveyed the risk preferences in 53 countries, significant cross-country differences in risk aversion depended on cultural factors. Their studies indicated that ambiguity aversion was much higher in Japan than in the USA, and Japanese people were slightly more risk averse when compared with global samples. Our results also possibly indicated that our subjects, all Japanese, were considerably risk-averse, and hence, their bids to buy assets were low, resulting in significantly low prices, which in turn led to lower expectations. However, our subjects would offer higher prices to sell assets if they were risk-averse.<sup>29</sup> Therefore, risk aversion is insufficient to explain the downward bias found in our experiment, so long as our subjects did not estimate the asset value only from the buyers' viewpoint. In addition, their inexperience in jar-guessing and other similar contests could also be a reason for imprecise estimates, but not their downward bias.

The reasons behind the negative bias in estimations could be a topic of interest for psychologists. Economists would be interested in the process of how the bias evolves in the market, rather than the bias in and of itself. Unfortunately, we did not discover any systematic effect of the market on prediction errors. Mean/median errors of predictions were erratic rather than systematic. In other words, collective predictions had no clear tendency to improve or worsen in our experimental markets.

However, we found that experience in the market reduced the diversity of predictions. Our subjects watched prices that converged at the median of submitted predictions. They adjusted their predictions toward the median prediction, even though they could not directly observe predictions of others. Hence, their predictions became similar as they participated in the same market. This can be a threat to *the collective intelligence* of the market. Surowiecki (2004) and Page (2007) insisted that individual

---

<sup>29</sup> Both selling and buying implied taking a risk in our experiment because transactions were settled by an unknown value. In so-called SSW asset markets, which originated from Smith et al. (1988), sellers avoided risks and buyers took risks since they traded a sort of lottery. Both sellers and buyers preferred a lower price in SSW markets if they were risk averse. In contrast, risk-averse sellers preferred a higher price in our markets, while a lower price was desirable for risk-averse buyers, when they took risks.

judgements should be diverse for *the wisdom of crowds* to emerge. Similarly, Treynor (1987) demonstrated the ways in which the shared error among investors deteriorated the accuracy of the market price.<sup>30</sup> Therefore, if the market reduces the diversity of individual expectations, it may destroy its own intelligence. Having said that, 24 traders in our experiment is too small to be called a crowd.<sup>31</sup> However, in light of the partial adjustment model of the prediction revision, we can guess that a larger laboratory market would reproduce the diminishing diversity of predictions, which may also hinder the emergence of *the collective intelligence*.

Our laboratory markets showed no intelligence in revealing the true value of the asset even during the early periods, when the subjects' predictions remained relatively heterogenous. Despite this, our market was efficient in disclosing a collective opinion as a cohort. Prices in our laboratory converged to the median of the subjects' predictions about the jar, i.e., to the equilibrium price that would balance demand and supply. It is well established that the price in the DA market converges to the equilibrium price. Demand-supply conditions that determine an equilibrium are given as parameters in traditional DA experiments that originated from Smith (1962). In contrast, the demand and supply in our laboratory were induced from the subjects' expectations about the jar, and these expectations changed during an experimental session. Despite such an insecure environment, the price converged to the equilibrium price. In conclusion, a failure in the market price to reveal the fundamental asset value was due to the preconceptions of the traders, and not the inefficiency of the market.

Conversely, if the initial median prediction had been precise, the market price would have revealed the true value of the asset. In view of the convergence of the price in our laboratory market, this revelation process would be robust to disturbances. Therefore, if subjects had enough practice to estimate the asset value, the market price would be a precise and stable estimator of the fundamental value. However, if we shocked the market, e.g., by changing the fundamental value at the stage when expectations of traders became homogenous, what would happen? Is a shock in the late stage more severe to the market stability than at the early stage with heterogenous expectation? We need further experiments to examine this topic.

## 5.2 Comparison to other studies

Smith et al. (1988), so-called SSW,<sup>32</sup> and Haruvy et al. (2007) studied the relation between the price formation and the expectation formation in experimental markets. Learning-to-forecast experiments, known as LtFEs,<sup>33</sup> also examined the relation,

<sup>30</sup> Subjects were cautioned after original guessing to allow for air space of the jar or the fact that the jar had thinner walls than a conventional jar. Treynor (1987) conjectured that such warnings caused shared error to creep into the estimates, and suggested that shared errors may be common in appraising companies too. He also insisted that shared errors created by published research are particularly important for asset prices in the real market.

<sup>31</sup> Our 48 subjects pooled from the two sessions can be a crowd since Treynor (1987) used 46 or 56 subjects in his classroom experiments.

<sup>32</sup> Many studies were derived from SSW. See, e.g., Palan (2013), Powell and Shestakova (2016), and Nuzzo and Morone (2017) for surveys.

<sup>33</sup> See Assenza et al. (2014) for a survey.

though their subjects concentrated on price forecasting since experimenters used a computer to make prices from submitted forecasts. In these experiments, subjects were required to forecast prices in the future. While our experiment also investigated the interdependence between price and expectation, the task of our subjects was to estimate the fundamental value of the asset, and not its price in the market. Other than this, our experiment differs from SSW's and LtFEs in many respects.

The life of the asset in our experiment was set for only one period unlike SSW's and LtFEs. Any inter-period trading which may bring capital gains was impossible in our experiment. This denied the profitability of the trend following behavior in which followers expect the future price to rise when the past prices increased. Even intra-period trading to secure a short-term profit was impossible, since our subjects could trade only once in a period and hence reverse contracts were excluded. They could neither resell an asset, nor buy it back from other subjects. Therefore, the uncertainty about the behavior of others, which is a cause for price deviations according to Akiyama et al. (2017), was irrelevant in our experimental market. No room for speculation in our market probably prevented bubbles, which often emerged in SSW's and LtFEs.<sup>34</sup>

Also, contrary to SSW's, the fundamental value of the asset was constant in our experiment. According to experiments by Noussair et al. (2001), constant fundamentals made bubbles less pronounced. Similarly, Kirchler et al. (2012) demonstrated that the declining fundamental value was the main driver for mispricing in experimental markets. We suspect that changing fundamentals could cause bubbles even in our non-speculative markets. An endogenously changing asset is a good prospect, since Soros (1987, p.57) said that a reflexive connection is interesting only if fundamentals are endogenously changeable.<sup>35</sup> What would happen if we changed the number of balls depending on the price in the market? We need further experiments to investigate this.

Even though no bubble occurred in our experiment, the fair price that reflected the fundamental value was never found in our markets because of a persisting downward bias in the subjects' expectations. This is not surprising because our subjects received no feedbacks about their errors and had no clues to correct estimates, which was distinct from SSW's and LtFEs. Real markets also lack the feedback about the fundamentals. Despite the feedback about errors in forecasting prices or profits of firms, or the information from the financial statements of firms, investors in the real world never know the true fundamental value. Our experiment was similar to the real world in this respect.

In SSW's and LtFEs, subjects knew the fundamentals since the probability distribution of dividends was revealed, even if they might not regard it fundamental. We determined the fundamental value by a jar with balls, not by a probability distribution. Our subjects probably regarded the jar as a fundamental since the transaction was settled at the value defined by the jar, but never knew the right value. Knight (1957)

---

<sup>34</sup> Our market was also free from the effect of the changing cash-to-asset value ratio, i.e., the ratio of all subjects' cash holdings and the total asset value, which is a factor to cause bubbles according to Kirchler et al. (2012). In our experiment, each subject was restricted to trading only one unit during a period by the rule, not by cash holdings.

<sup>35</sup> Soros (1987, p.57) said "It is possible to have a reflexive connection between stock prices and the prevailing bias even if the fundamentals remain unaffected, but the connection becomes interesting only if the fundamentals are also involved."

and Keynes (1921) contemplated the uncertainty that has no mathematical probability. Further, Ellsberg (1961) experimentally discovered the effect of the ambiguity, which was not specified by the probability distribution. Although the value of our experimental asset was deterministic, our subjects never knew its fundamental value. It was ambiguous for the subjects not because of the uncertainty due to the mathematical probability, but because of the limitation of their cognitive ability.

To extract the effect of this ambiguity, we need additional experiments. For example, we may introduce the uncertainty of the fundamental value through the probability distribution determined by the distribution of submitted predictions about the jar. What if subjects participated in both markets: the market of the ambiguous asset and the market of the asset of measurable uncertainty?

### 5.3 Future issues

We need more work to examine the exact mechanism of price convergence. In our experiment, approximately half of the buyers with higher estimations of the asset value were bidding up—toward their estimations. The other half were asking to sell, conceding their order prices toward their lower estimated values. Such a process might necessarily make the price to converge at the median of estimations. However, the following questions arise: Why was the number of buyers approximately same as the number of sellers? Why did order prices of each side meet at the middle of the estimated values? Were there no inequalities in bargaining power? We need to scrutinize our experimental data more elaborately to answer these questions. Additional experiments might be necessary.

We studied our subjects' expectation revision process with a simple linear model in Sect. 4. The disturbance term in this linear model had a positive correlation with the subjects' expected value (see Table 13), suggesting some nonlinearity in the nature of the revision process. A non-linear model is probably more appropriate to represent the behavior of our subjects. However, nine periods in our experiment were too short for examining individual behavior. It is necessary to conduct a more-than-12-period session.

We modeled our subjects' expectation formation as the partial adjustment process to the median of submitted predictions. Even though the subjects had no direct information about predictions submitted by others, this model provided a good approximation of the subjects' behavior, which was not surprising since prices in markets tended to converge to the median of predictions. Nevertheless, targets of adjustment probably differ among subjects, depending on their focus about what to observe in the market. The variety of the target, represented by  $\epsilon_{i,t}$  in our model (7), is part of the driving force to correct or exacerbate a prediction bias. Eye tracking of subjects would reveal this variety. Interviews and questionnaires to subjects after the experimental sessions would also be useful.

## 6 Concluding remarks

In our experiment, the subjects' predictions about the number of balls in a jar had a downward bias. Transaction prices of the asset whose value was defined by the jar also showed a downward bias. Prices reflected the downward bias of the subjects' expectations about the fundamental value because of their auctioning behavior. Subjects who predicted higher values tended to bid. Lower value predictors were apt to ask. Consequently, the prices converged at the median prediction, which had a downward bias in our experiment. We need additional studies to reveal the precise mechanism of this convergence to the equilibrium price defined by the median prediction.

As the subjects alternately repeated jar-guessing and asset trading, their expectations about the fundamental value changed. Although our market neither improved nor deteriorated the accuracy of predictions submitted by the subjects, it made them less heterogenous. We examined the prediction revision of our subjects using the partial adjustment model, which targets the median of submitted predictions. Despite our experimental design that did not allow the subjects to observe predictions of others, this model was useful to explain experimental outcomes, probably because prices in the market reflected the prediction median. Further investigation is necessary to uncover the exact target of the adjustment process, which may vary among subjects.

If all subjects strictly followed the partial adjustment process with the target of the median prediction, the central tendency of their predictions would almost stay at the initial region. Therefore, some deviation from the partial adjustment is inevitable not only to disturb precise predictions, but also to correct a bias in initial expectations. Such deviations possibly come from over-adjusting behavior, perturbations in the market, individual variation in the target, and personal uncertainty.

**Acknowledgements** We are grateful to Dr. Xi Fu of University of Liverpool for an interesting discussion at the 5th International Symposium in Computational Economics and Finance. We appreciate Prof. Sobei H. Oda and Prof. Yoshio Iida of Kyoto Sangyo University for their kind advice and support for conducting experiments. We also thank anonymous referees of the Journal of Behavioral and Experimental Finance, who reviewed and commented on our earlier manuscript, from which this paper originates. We owe a great deal to anonymous reviewers of the Journal of Economic Interaction and Coordination in revising our manuscript. All possible errors are our own.

## Declaration

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Table 17** Statistical summary of predictions in the two sessions

Period <i>t</i>	1	2	3	4	5	6	7	8	9
Mean	309.3	280.9	288.1	284.5	285.0	286.8	284.5	284.7	277.6
Median	300.0	280.0	280.5	267.5	267.5	280.0	289.5	280.0	277.5
Standard deviation	132.2	98.5	109.7	105.9	103.0	109.2	97.9	104.6	100.3
Interquartile range	207.5	135.0	126.3	102.5	110.0	130.0	105.5	105.0	103.3
Maximum	614	500	600	600	550	650	555	600	500
Minimum	50	60	65	70	70	65	60	1	1
Range	564	440	535	530	480	585	495	599	499
Skewness	0.208	-0.057	0.347	0.730	0.565	0.712	0.402	0.154	-0.207
Kurtosis	-0.625	-0.338	0.484	0.728	0.289	1.507	0.698	1.484	0.633
Mean prediction error	-74.7	-103.1	-95.9	-99.5	-99.0	-97.3	-99.5	-99.3	-106.4
<i>T</i> value	-3.91	-7.25	-6.06	-6.51	-6.66	-6.17	-7.04	-6.58	-7.35
<i>P</i> value	1.5E-04	1.7E-09	1.1E-07	2.3E-08	1.3E-08	7.5E-08	3.5E-09	1.8E-08	1.2E-09

48 observations for each period

## Appendix

### A Statistics about predictions

The table above shows statistics about predictions submitted by the subjects at each period in the two sessions. Each period involves 48 observations (Table 17).

*T* value above is the value of the t-test statistic to test the null hypothesis that the mean prediction error is non-negative. The calculation formula is the next.

$$\frac{\bar{V}_t - 384}{S_t/\sqrt{48}}$$

where  $\bar{V}_t = \frac{1}{48} \sum_{k=1}^2 \sum_{i=1}^{24} V_{i,t,k}^e$  and  $S_t = \sqrt{\frac{1}{47} \sum_{k=1}^2 \sum_{i=1}^{24} (V_{i,t,k}^e - \bar{V}_t)^2}$ .

*P* value is the single-sided significance probability which is defined as follows.

$$\Pr[\text{Test Statistic} < T \text{ value}]$$

Skewness is calculated by the next formula.

$$\frac{48^2}{47 \times 46} \frac{\frac{1}{48} \sum_{k=1}^2 \sum_{i=1}^{24} (V_{i,t,k}^e - \bar{V}_t)^3}{S_t^3}$$

Kurtosis is calculated as follows.

$$\frac{49 \times 48^2}{47 \times 46 \times 45} \frac{\frac{1}{48} \sum_{k=1}^2 \sum_{i=1}^{24} (V_{i,t,k}^e - \bar{V}_t)^4}{S_t^4} - 3 \cdot \frac{47^2}{46 \times 45}$$



**Table 18** Frequency of predictions in the two sessions

Period $t$	1st class ( $-\infty, 118$ )	2nd class [118, 194)	3rd class [194, 270)	4th class [270, 346)	5th class [346, 422)	6th class [422, 498)	7th class [498, 574)	8th class [574, $+\infty$ )
1	2	10	7	<u>11</u> *	7	5	5	1
2	2	7	11	<u>14</u> *	11	2	1	0
3	2	8	9	<u>14</u> *	12	0	2	1
4	1	8	<b>15</b> *	<u>14</u>	4	4	1	1
5	1	7	<b>16</b> *	<u>14</u>	5	2	3	0
6	1	10	10	<u>14</u> *	9	2	1	1
7	1	8	9	<u>20</u> *	6	2	2	0
8	2	7	9	<u>19</u> *	7	3	0	1
9	2	7	10	<u>20</u> *	5	3	1	0

The asterisk \* is attached to the frequency of the class to which the median belongs

The underlined number is the frequency of the class to which the mean belongs

The bold letter indicates the modal frequency

The total frequency is 48 for each period

## B Frequency table of predictions

The table above represents the frequency distribution of the predictions submitted in Sessions 1 and 2. Figure 2 in Sect. 3.1.1 graphically depicts the first period of this table. Predictions at  $t = 4$  were submitted right after the second jar observation. Those submitted at  $t = 7$  directly followed the third observation (Table 18).

## C Individual regression of subjects' prediction revision

See Tables 19 and 20.

**Table 19** Individual regression of Session 1

Subject	Intercept			Regression coefficient			$\sqrt{R^2}$
	$\hat{\alpha}_i$	<i>T</i> value	<i>P</i> value	$\hat{\beta}_i$	<i>T</i> value	<i>P</i> value	
1	40.069*	2.450	0.050	0.989*	6.819	0.000	0.941
2	162.669*	10.862	0.000	1.617*	11.752	0.000	0.979
3	19.335	0.675	0.525	-0.273	-0.605	0.567	0.240
4	79.227	2.077	0.083	0.602	2.284	0.062	0.682
5	6.648	0.731	0.493	0.071	0.480	0.648	0.192
6	-24.949*	-2.597	0.041	1.237*	2.971	0.025	0.772
7	-118.140*	-14.710	0.000	1.096*	15.951	0.000	0.988
8	5.078	0.632	0.550	0.135	0.426	0.685	0.171
9	89.269	1.670	0.146	0.530	1.791	0.123	0.590
10	-55.707	-1.165	0.288	0.367	1.335	0.230	0.479
11	-32.322	-0.756	0.478	0.081	0.270	0.796	0.109
12	-12.160	-0.912	0.397	0.655*	2.986	0.024	0.773
13	16.340	0.649	0.540	0.635	1.681	0.144	0.566

**Table 19** continued

Subject	Intercept			Regression coefficient			$\sqrt{R^2}$
	$\hat{\alpha}_i$	<i>T</i> value	<i>P</i> value	$\hat{\beta}_i$	<i>T</i> value	<i>P</i> value	
14	8.000	0.732	0.492	1.030*	2.663	0.037	0.736
15	-0.526	-0.053	0.960	0.064	0.635	0.549	0.251
16	-45.369*	-2.826	0.030	0.709*	3.213	0.018	0.795
17	3.438	0.196	0.851	0.433	1.649	0.150	0.558
18	113.204	2.188	0.071	1.187*	2.822	0.030	0.755
19	0.000	NA	NA	0.000	NA	NA	1.000
20	13.345	1.358	0.223	0.454	1.842	0.115	0.601
21	9.366	0.341	0.745	1.148*	3.144	0.020	0.789
22	9.519	0.461	0.661	-0.139	-0.420	0.689	0.169
23	22.794*	2.591	0.041	0.817*	3.147	0.020	0.789
24	57.655	1.263	0.254	0.596	1.465	0.193	0.513

Estimates significantly different from 0 at 5% level are denoted by \*

*P* value indicates a double-sided significance probability

Subject 19 never changed predictions

The number of observations is 8 for each regression

### D Asymmetry of prediction distribution

Figure 1 shows that the discrepancy between the mean and median of predictions increased around the middle periods, implying that the prediction distribution would be more asymmetric in these periods. The values of the skewness of the prediction distribution were the largest in the middle three periods, as shown in Fig. 15. Although

**Table 20** Individual Regression of Session 2

Subject	Intercept			Regression coefficient			$\sqrt{R^2}$
	$\hat{\alpha}_i$	<i>T</i> value	<i>P</i> value	$\hat{\beta}_i$	<i>T</i> value	<i>P</i> value	
1	26.609	0.611	0.564	0.587	1.567	0.168	0.539
2	-8.451	-2.366	0.056	0.287	1.871	0.111	0.607
3	296.071*	2.988	0.024	1.375*	3.321	0.016	0.805
4	-19.635	-0.750	0.482	0.532	1.527	0.178	0.529
5	0.856	0.092	0.930	0.282	1.360	0.223	0.485
6	-45.534	-1.657	0.149	0.393	1.793	0.123	0.591
7	-107.735*	-3.961	0.007	1.123*	4.050	0.007	0.856
8	-16.847	-1.202	0.275	0.211	1.306	0.239	0.471
9	9.175	0.889	0.408	0.636	1.890	0.108	0.611
10	-5.802	-1.318	0.236	0.207	1.134	0.300	0.420
11	0.000	NA	NA	0.000	NA	NA	1.000
12	15.625	1.304	0.240	0.358	1.457	0.195	0.511
13	5.532	0.965	0.372	-0.798	-1.728	0.135	0.576
14	16.893	0.506	0.631	0.170	0.624	0.556	0.247
15	25.063	1.342	0.228	0.797*	2.740	0.034	0.745

**Table 20** continued

Subject	Intercept			Regression coefficient			$\sqrt{R^2}$
	$\hat{\alpha}_i$	<i>T</i> value	<i>P</i> value	$\hat{\beta}_i$	<i>T</i> value	<i>P</i> value	
16	-71.485	-1.835	0.116	0.676	1.796	0.123	0.591
17	-62.719*	-3.121	0.021	0.326*	3.224	0.018	0.796
18	6.288	0.456	0.664	0.134	0.650	0.540	0.256
19	-23.464	-0.786	0.462	0.523	1.454	0.196	0.510
20	-31.212	-0.972	0.369	0.327	1.178	0.283	0.433
21	-0.704	-0.181	0.862	0.104	0.890	0.408	0.341
22	13.475	1.288	0.245	0.288	1.240	0.261	0.452
23	-10.465	-0.238	0.819	0.187	0.432	0.681	0.174
24	212.902	1.955	0.098	0.850	2.139	0.076	0.658

Estimates significantly different from 0 at 5% level are denoted by \*

*P* value indicates a double-sided significance probability

Subject 11 never changed predictions

The number of observations is 8 for each regression

additional experiments are required to confirm whether the distribution has a tendency to become the most asymmetric in the middle periods of a session, the following subsections examine the reason for the variation pattern of the asymmetry found in our two-session experiment.

**Table 21** Summary of prediction frequency

Period $t$	Left-end class ( $-\infty, 194$ )	Middle class [270, 346)	Right-end class [422, $+\infty$ )
1	12	11	11
2	9	14	3
3	10	14	3
4	9	14	6
5	8	14	5
6	11	14	4
7	9	20	4
8	9	19	4
9	9	20	4

**Table 22** Average of the predictions in the right-end class

Period $t$	1	2	3	4	5	6	7	8	9
Avg. prediction	494.5	460.0	533.3	486.7	494.0	517.5	493.8	486.3	467.5

### D.1 Prediction behavior behind the fluctuation of the skewness

Table 21 is an excerpt of the frequency table in Appendix B. The middle class is identical to the fourth class of Table 18, which includes the mean prediction at every period. The right-end class summarizes the sixth, seventh, and eighth classes of Table 18 that are more than one class above the fourth class. The left-end class is a summary of classes that are more than one class below the fourth class: the first and second classes of Table 18.

The frequency of the right-end class doubled at  $t = 4$ . Two subjects in the fourth class and one subject in the lower end of the fifth class at  $t = 3$  entered the right-end class at  $t = 4$ , whereas three in the right-end class at  $t = 3$  remained in this class at  $t = 4$ . This rightward swing of predictions increased the skewness to the maximum. The right-end class kept the frequency of 4 after  $t = 6$ . The same three subjects remained in this class after  $t = 4$ . Table 22 shows that the average of the predictions in the right-end class declined monotonically after  $t = 6$ . The skewness also continued to decrease after  $t = 6$ .

The attracting force of the market price likely led the average of the predictions in the right-end class to decline as transaction prices were always below the level of this class. The upward swing of three predictions into the right-end class at  $t = 4$ , which seemed to cause the skewness to surge, occurred directly after the reobservation of the jar. The next subsection examines the impact of the jar observation.

### D.2 Effect of jar observation

Every subject observed a jar three times. The first observation was before  $t = 1$ . The second observation was right after  $t = 3$  and before  $t = 4$ . The third observation period was between  $t = 6$  and  $t = 7$ .

Table 23 represents how actively our subjects revised their predictions. The number of zero-revisions reflects how many subjects submitted the same predictions as the previous predictions: the number of  $i$ 's that satisfied  $\Delta V_{i,t,k}^e = 0$ . The average absolute revision is the average of the absolute change of predictions:  $\sum_{k=1}^2 \sum_{i=1}^{24} |\Delta V_{i,t,k}^e|/48$ . The revision at  $t$ ,  $\Delta V_{i,t,k}^e$ , is a change from  $t$  to  $t + 1$ :  $\Delta V_{i,t,k}^e = V_{i,t+1,k}^e - V_{i,t,k}^e$ . Therefore, revisions at  $t = 3$  and  $t = 6$  directly followed the jar observation.

The number of zero-revisions had double bottoms at  $t = 3$  and  $t = 6$ . The average absolute revision also had local peaks at  $t = 3$  and  $t = 6$ . The peak at  $t = 3$  is the largest. It seems that the jar observation encouraged subjects to actively revise their predictions, though the second reobservation before the sixth revision had less impact on activities related to revising predictions.

The first reobservation before the third revision induced both upper and lower revisions substantially. Table 24 reports the aggregate revision in each direction. The aggregate upward revision is the sum of positive prediction changes:  $\sum_{k=1}^2 \sum_{i=1}^{24} \max\{\Delta V_{i,t,k}^e, 0\}$ . We sum the negative prediction changes to obtain the aggregate downward revision:  $\sum_{k,i} \min\{\Delta V_{i,t,k}^e, 0\}$ .

The significantly positive and negative aggregate revisions at  $t = 3$  imply that the subjects revised their predictions drastically in both directions after the first reobservation. The predictions moved rightward and leftward on a large scale. These bidirectional extreme movements resulted in a distinctively skewed distribution of the predictions at  $t = 4$ .

The reobservation of the jar had an impact on predictions. Nevertheless, the second reobservation before the sixth revision affected them less prominently. Moreover, such inspections of the fundamentals by a jar-observation did not necessarily improve predictions of the fundamental value. Table 25 and Fig. 16 present the aggregate absolute error of predictions:  $\sum_{i=1}^{24} |V_{i,t,k}^e - 384|$  for  $k = 1$  and 2, and  $\sum_{k=1}^2 \sum_{i=1}^{24} |V_{i,t,k}^e - 384|$  for the pooled sample of the two sessions. It is clear the first reobservation rendered predictions at  $t = 4$  inaccurate. The effect of the second reobservation is slight and indefinite. It improved predictions at  $t = 7$  for Session 2 but caused them to worsen for Session 1. Furthermore, it slightly alleviated the aggregate error for the pooled sample of the two sessions.

**Table 23** Prediction revisions in the two sessions

Period $t$	1	2	3	4	5	6	7	8
# of Zero-revisions	19	24	10	24	19	11	20	23
Avg. absolute revision	40.6	32.1	49.0	10.9	17.1	27.2	19.7	16.6

The revisions at  $t = 3$  and  $t = 6$  directly followed the jar observation

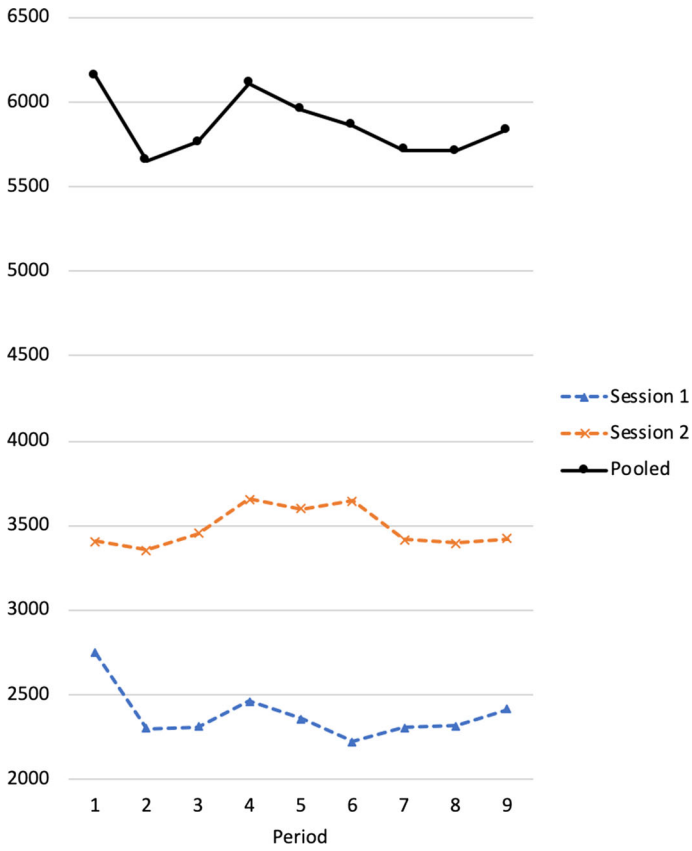


Fig. 16 Aggregate absolute errors of predictions

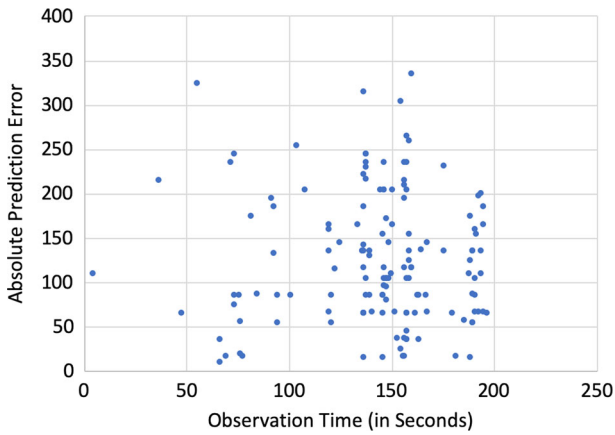


Fig. 17 Observation time and absolute prediction error

**Table 24** Aggregate prediction revision in the two sessions

Period $t$	1	2	3	4	5	6	7	8
Upward revision	293	942	1088	275	452	598	478	230
Downward revision	-1655	-598	-1262	-250	-368	-707	-469	-568

The revisions at  $t = 3$  and  $t = 6$  directly followed the jar observation

**Table 25** Aggregate absolute errors of predictions

Period $t$	1	2	3	4	5	6	7	8	9
Session 1	2747	2297	2310	2459	2357	2219	2303	2317	2415
Session 2	3406	3356	3453	3654	3599	3645	3416	3397	3423
Pooled	6153	5653	5763	6113	5956	5864	5719	5714	5838

The first reobservation of the jar was between  $t = 3$  and  $t = 4$

The second reobservation of the jar was between  $t = 6$  and  $t = 7$

## E Observation time and prediction accuracy

The time spent on the jar observation was not correlated with the accuracy of the predictions. In each session, every subject observed a jar three times; thus, we have a sample of 144 observations as we conducted two sessions and 24 subjects participated in each session. The absolute error of predictions immediately following the jar observation, that is,  $|V_{i,t,k}^e - 384|$  at  $t = 1, 4, 7$ , showed no correlation with these 144 observation times as the correlation was only  $-0.027$  (see the scatter diagram in Fig. 17).

However, a sample of extreme observations demonstrated a negative correlation, suggesting that very long observations entailed small prediction errors. If we choose the longest and shortest times from every observation period in each session, we obtain 12 observations. These extreme observation times were negatively correlated with the absolute prediction errors, showing a  $-0.44$  correlation (see Fig. 18). However, we can detect such correlation only in very extreme observations. If we choose the two longest and two shortest times from each period,<sup>36</sup> these 24 observation times show no correlation with the absolute prediction error, with a  $-0.059$  correlation.

A long observation time did not necessarily indicate elaborate observation. Subjects would require a longer duration of time if their observation process was awkward. In fact, the 12 observations accounting for the six longest and six shortest times among the 144 total observations were not correlated with absolute prediction errors: the correlation was  $-0.068$ . The longest six were all during the first observation period, while the shortest six were during the last period. Therefore, these selected observation times reflected the subjects' experience in the jar observation.

The relation between the observation time and the prediction error in our experiment was not straightforward, being entangled with various factors.

<sup>36</sup> If there were ties among observation times, we selected times with a smaller absolute prediction error for the longest time sample and times with a larger absolute error for the shortest sample.

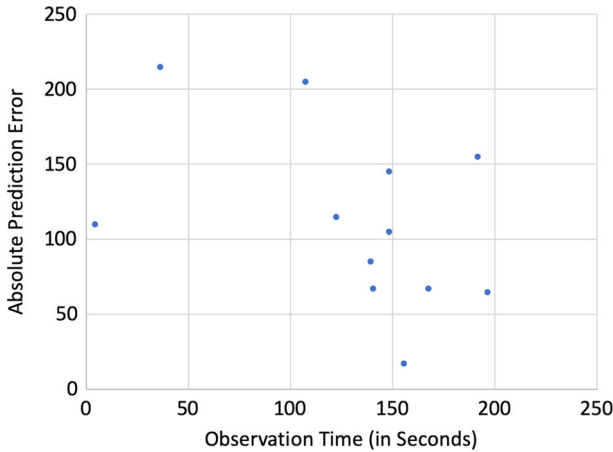


Fig. 18 Extreme observation time and absolute prediction error

## F Comparison of sessions

Table 25 and Fig. 16 show that the subjects in Session 1 as a cohort predicted the fundamental value better than those in Session 2. Figures 9 and 10 demonstrate that the price reflected this difference of prediction accuracy as the absolute deviation from the fundamental value was smaller in Session 1 than in Session 2. The subjects in Session 1 tended to take more time than those in Session 2 to observe a jar at each period except the third period, as shown in Table 1. However, a longer jar observation did not necessarily imply more precise estimation of the fundamental value, as we argue in Appendix E. Therefore, longer observations in the first and second periods were insufficient to explain the reason why the predictions were relatively accurate in Session 1.

However, observation time declined sharply at the third observation period in Session 1. It is possible that the subjects in Session 1 were so confident about their estimation that they did not need more time to observe the jar during the third period. They seemed to be quick to become confident compared to the subjects in Session 2. Figs. 9 and 10 also indicate that the price was more convergent to the equilibrium in Session 1 than in Session 2. The standard deviation and interquartile range of predictions were often smaller in Session 1 than in Session 2, as shown in Table 26. Table 27 indicates that the subjects in Session 1 submitted more bids/asks<sup>37</sup> than the subjects in Session 2, especially in later periods. Such differences in unanimity regarding the fundamental value and auctioning activities in markets seemed to cause the price convergence to differ.

Although we are not certain about subjects' traits that caused such differences, we noticed distinctive characteristics of the subjects in Session 1. The first author, an experimenter, described in a log-book that the subjects in Session 1 seemed to be quieter than usual. The second author, a laboratory assistant, also reported his

<sup>37</sup> Taking a price on the order book is included in the number of bids and asks.



**Table 26** Standard deviation and interquartile range of predictions

Period $t$		1	2	3	4	5	6	7	8	9
SD	Session 1	132.2	92.6	97.3	90.4	89.0	81.8	84.6	101.8	100.5
	Session 2	122.3	97.6	117.1	116.0	110.4	129.4	108.1	106.1	97.5
IQ range	Session 1	200	117.25	120	132.5	121.25	105	102.25	116.25	83.75
	Session 2	168.75	132	136.25	129.25	132.5	170	105	104.25	120.75

**Table 27** Number of submitted bids and asks

Period $t$	1	2	3	4	5	6	7	8	9
Session 1	30	43	27	46	39	48	47	41	40
Session 2	30	38	46	36	45	35	42	41	36

**Table 28** Subjects' attributes

		Session 1	Session 2
School year	Freshman	6	6
	Sophomore	7	7
	Junior	6	5
	Senior	5	6
Gender	Male	16	15
	Female	8	9
Academic major	Economics	6	7
	Business administration	3	2
	Law	3	3
	Sociology	3	1
	Foreign studies	0	3
	Cultural studies	4	2
	Science	2	2
	Information science	1	3
	Life science	2	1

Each of the three categories reflects the 24 subjects in each session

impression that the subjects were calm, noting that none of them failed to correctly fill out receipts for their cash rewards. However, the school year, gender, and academic major of the subjects were similar between the two sessions. Table 28 displays how many subjects had each attribute. We likely needed to conduct a personality test or cognitive reflection test to identify the traits of the subjects that caused the experimental outcomes to differ.<sup>38</sup>

## References

Akiyama Eizo, Hanaki Nobuyuki, Ishikawa Ryuichiro (2017) It is not just confusion! strategic uncertainty in an experimental asset market. *Econ J* 127:F563–F580

<sup>38</sup> Akiyama et al. (2017) used the scores of the subjects in a cognitive reflection test to reveal the difference in the effect of strategic uncertainty on mispricing.

- Assenza T, Bao T, Hommes C, Massaro D (2014) Experiments on expectations in macroeconomics and finance. Experiments in macroeconomics. Emerald Group Publishing Limited, Bingley, pp 11–70
- Bao T, Hommes C, Sonnemans J, Tuinstra J (2012) Individual expectations, limited rationality and aggregate outcomes. *J Econ Dyn Control* 36:1101–1120
- Bao T, Hommes CH, Makarewicz TA (2017) Bubble formation and (in)efficient markets in learning-to-forecast and optimise experiments. *Econ J* 127(605):F581–F609
- Carhart Mark M (1977) On Persistence in Mutual Fund Performance. *J Finance* 52:57–82
- Colasante A, Alfarano S, Camacho-Cuena E (2019) The term structure of cross-sectional dispersion of expectations in a learning-to-forecast experiment. *J Econ Interact Coordin*. <https://doi.org/10.1007/s11403-019-00245-6>
- Cowles 3rd, Alfred (1933) Can stock market forecasters forecast? *Econometrica* 1:309–324
- Ellsberg Daniel (1961) Risk, ambiguity, and the savage axioms. *Quart J Econ* 75:643–669
- Fama Eugene F (1970) Efficient capital markets: a review of theory and empirical work. *J Finance* 25:383–417
- Fischbacher Urs (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Exp Econ* 10:171–178
- Friedman Daniel, Sunder Shyam (1994) Experimental methods: a primer for economists. Cambridge University Press, New York
- Galton Francis (1907) Vox populi. *Nature* 75:450–451
- Haruvy E, Lahav Y, Noussair CN (2007) Traders' expectations in asset markets: experimental evidence. *Am Econ Rev* 97:1901–20
- Heemeijer P, Hommes C, Sonnemans J, Tuinstra J (2009) Price stability and volatility in markets with positive and negative expectations feedback: an experimental investigation. *J Econ Dyn Control* 33:1052–1072
- Keynes John M (1921) A treatise on probability. Macmillan and Co., London
- Kirchler M, Huber J, Stöckl T (2012) Thar she bursts: reducing confusion reduces bubbles. *Am Econ Rev* 102:865–883
- Knight Frank H (1957) Risk, uncertainty and profit. Houghton Mifflin Co., Boston and New York
- Lorge I, Fox D, Davitz J, Brenner M (1958) A survey of studies contrasting the quality of group performance and individual performance, 1920–1957. *Psychol Bull* 55:337–372
- Maloney MT, Mulherin JH (2003) The complexity of price discovery in an efficient market: the stock market reaction to the Challenger crash. *J Corp Finan* 9:453–479
- Marimon R, Spear SE, Sunder S (1993) Expectationally driven market volatility: an experimental study. *J Econ Theory* 61:74–103
- Noussair C, Robin S, Ruffieux B (2001) Price bubbles in laboratory asset markets with constant fundamental values. *Exp Econ* 4:87–105
- Nuzzo Simone, Morone Andrea (2017) Asset markets in the lab: a literature review. *J Behav Exp Financ* 13:42–50
- Page Scott E (2007) The difference: how the power of diversity creates better groups, firms, schools, and societies. Princeton University Press, New Jersey
- Palan Stefan (2013) A Review of bubbles and crashes in experimental asset markets. *J Econ Surv* 27:570–588
- Powell Owen, Shestakova Natalia (2016) Experimental asset markets: a survey of recent developments. *J Behav Exp Financ* 12:14–22
- Rieger MarcO, Wang M, Hens T (2014) Risk preferences around the world. *Management Science*, Published online in Articles in Advance 21 Feb 2014
- Smith Vernon L (1962) An experimental study of competitive market behavior. *J Polit Econ* 70:111–137
- Smith Vernon L, Suchanek Gerry L, Williams Arlington W (1988) Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica* 56:1119–1151
- Soros George (1987) Alchemy of finance. Wiley, New Jersey
- Surowiecki James (2004) The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations. Anchor Books, New York
- Treynor Jack L (1987) Market efficiency and the bean jar experiment. *Financ Anal J* 43:50–53
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185:1124–1131