ISSN (Print) 0473-453X
Discussion Paper No. 1213
ISSN (Online) 2435-0982

# AN EXPERIMENT ON A MULTI-PERIOD BEAUTY CONTEST GAME

Nobuyuki Hanaki
Yuta Takahashi

Revised January 2024
September 2023

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

# An Experiment on a Multi-Period Beauty Contest Game*

Nobuyuki Hanaki† Yuta Takahashi‡

January 23, 2024

**Abstract**

We present and conduct a novel experiment on a multi-period beauty contest game motivated by the canonical New-Keynesian model. Participants continuously provide forecasts for prices spanning multiple future periods. These forecasts determine the price for the current period and participants' payoffs. Our findings are threefold. First, the observed prices in the experiment deviate more from the rational expectations equilibrium prices under strategic complementarity than under strategic substitution. Second, participants' expectations respond to announcements of future shocks on average. Finally, participants employ heuristics in their forecasting; however, the choice of heuristic varies with the degree of strategic complementarity.

**Keywords:** Expectation formation, Learning-to-forecast experiment

**JEL Code:** C92, D84, E70

*We thank Elena Asparouhova, Te Bao, John Duffy, Charles Noussair, Yoshiyasu Ono, Lube Petersen, Marc Willinger, and the participants at 2023 CEFM workshop (Nagasaki), AP-ESA 2023 (Seoul), EF 2023 (Sofia), and 2023 ESA world meeting (Lyon) for their comments and helpful discussion. This research has benefited from the financial support of (a) the Joint Usage/Research Center (JJURC), the Institute of Social and Economic Research (ISER), and Osaka University, and (b) Grants-in-aid for Scientific Research Nos. 18K19954, 20H05631, and 23H00055 from the Japan Society for the Promotion of Science. The design of the experiment reported in this paper was approved by the IRB of ISER (#20221102) in November 2022. The experiment is preregistered at aspredicted.org (#128898), which is available at https://aspredicted.org/ix6th.pdf. We thank Yuka Tamura and Yuta Shimodaira for their help in preparing software for the experiment, and Yuki Hamada, Hiroko Shibata, Saori Shikibu, and Satsuki Yamada for their support in conducting the experiment.

†Corresponding author. Institute of Social and Economic Research, Osaka University, and University of Limassol. 6-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp. TEL: +81 (6) 6879 8552 FAX: +81 (6) 6879 8583

‡Institute of Economic Research, Hitotsubashi University

# 1 Introduction

The assumption that individuals form rational expectations is pervasive and often taken for granted in macroeconomic modeling. Over the last five decades, a substantial body of macro and experimental economics literature has evaluated the rational expectations equilibrium (REE) hypothesis. One notable finding from the macroeconomics literature is that models with the strong form of rational expectations can lead to erratic implications, such as the forward-guidance puzzle. Consequently, researchers have become increasingly skeptical of this assumption, proposing alternative frameworks (for example, Angeletos and Lian, 2018; Farhi and Werning, 2019; García-Schmidt and Woodford, 2019).

Another finding emerges from the literature on so called "learning-to-forecast" experiments (Marimon and Sunder, 1993). This literature finds that the results of numerous experiments are inconsistent with the strong form of rational expectations (see, among others, Hommes et al., 2005; Heemeijer et al., 2009; Bao et al., 2017). In these experiments, unlike an experiment where participants make their forecasts about exogenously-generated series (e.g., Afrouzi et al., 2023), participants submit an expectation about an aggregate outcome (e.g., prices), and the aggregate outcome is endogenously determined through an aggregation of these individual forecasts.[1]

[1] Note that unlike experiments in which participants actually trade in the market (see, e.g., Smith et al., 1988; Haruvy et al., 2007; Akiyama et al., 2017; Asparouhova et al., 2016; Crockett et al., 2019), set prices (see, e.g., Fehr and Tyran, 2008; Noussair et al., 2015; Orland and Roos, 2013, 2019; Petersen, 2015) or quantities (see, e.g., Bosch-Domènech and Vriend, 2003; Huck et al., 1999, 2004; Offerman et al., 2002), in learning-to-forecast experiments, participants do not trade, set prices or quantities, but only forecast. This eliminates the need for participants to trade or set prices (or quantity) optimally and thus allows us to focus on their expectation formation and its aggregate consequences. See, Bao et al. (2013), for a comparison with learning-to-forecast experiments with those where participants need to decide on the quantities (which the authors call "learning-to-optimize" experiments). The learning-to-forecast experimental framework has been used to investigate policy-relevant questions, such as the causal impacts of the central bank communication on the expectation and the aggregate outcomes (see, e.g., Kryvtsov and Petersen, 2021; Mokhtarzadeh and Petersen, 2021).

This feedback effect is often referred to as an *expectational feedback*: past events can shape individual expectations which, in turn, affect the current aggregate market outcome. As a result, the results of these experiments can deviate substantially, depending on the nature and the strength of the feedback, from those predicted by the rational expectations hypothesis. Instead, they can be better explained by models with backward-looking expectation formation (see, for example, Anufriev and Hommes, 2012; Evans et al., 2021).

Notably, most of these existing learning-to-forecast experiments are essentially single-period.[2] Participants only need to forecast the outcome of today or in one future period (e.g., tomorrow), and the realized outcome is a function of submitted forecasts regarding this single-period (see Bao et al., 2021, for a survey).[3] As a result, it is difficult to provide useful insights for macroeconomists who study expectation formation in a multi-period environment. In a typical macroeconomic model and more generally a dynamic game, individuals forecast not only actions of others in one period, but also their actions in multiple future periods. And these forecasts about multiple future periods determine the current outcome. The single-period framework does not enable us to analyze how these series of expectations change in such a multi-period environment, for example, in response to news about future shocks.

We contribute to the literature by filling this gap. We propose and conduct a novel experiment using a multi-period beauty contest game motivated by the core part of the canonical New-Keynesian model (e.g., Woodford (2003) and Galí (2015), which is initially developed by Calvo (1983)). In our theoretical model, there are many firms, and each firm faces the same linear demand curve. The demand for this firm depends

[2]There are several studies, mostly recent, who study multi-period experiments. We discuss them in detail later.

[3]This approach stands in stark contrast to survey-based studies on expectation formation, which elicit expectations of the same future periods multiple times and study their evolution (see, e.g., Coibion et al., 2018).

on the price of the firm and the aggregate price. The demand decreases as the firm increases its price, but can increase or decrease as other firms increase their prices. If firms can choose their prices freely, firms choose them in order to maximize their profit given the prices of others. A REE price, in this case, is a fixed point of this process and often referred to as the optimal flexible price.

Now we introduce friction that prevents firms from choosing this flexible price motivated by Calvo (1983). We assume that they can only change their prices with a probability of $1 - \theta$. This friction compels them to take future prices into account since they affect profit. The market has this additional dynamic feedback effect since expectations of future prices directly affect today's action.

We transform our theoretical model into a multi-period version of the "learning-to-forecast" game. In each period, participants submit their forecasts for prices in multiple future periods. Each of these submitted forecasts can determine participants' payoff as well as prices.

We undertake a 3 by 2 between-subjects experimental design. Our first focus is on games that feature strategic complementarity, as well as weak and strong substitution. Furthermore, during the course of the experiment, we introduce two upward shifts in the demand curve, which we refer to as "shocks." Our second focus is on when these shocks occur unexpectedly, as well as on cases where shocks are announced to participants two periods in advance.

Using this experimental framework, we begin by testing two hypotheses. In the first hypothesis, we investigate the effect of the strategic environment (Hanaki et al., 2019). We test whether observed prices deviate more from the REE prices under strategic complementarity than under strategic substitution. This hypothesis test is mainly to confirm existing results of past learning-to-forecast experiments within our new experimental framework.[4] In the second hypothesis, we examine a nature of forward-lookingness of expectation. Specifically, we assess whether participants adjust their expectations for future prices in response to announcements of future shocks in environments where prices move along the REE in periods before the announcement. Our proposed multi-period framework facilitates the testing of this latter hypothesis, a feat unachievable with single-period experimental designs. We show that our experimental results validate the two hypotheses.

We then proceed to analyze how participants form their expectation in our multi-period beauty contest game. We uncover various facts. First, as observed in single-period experiments, a sizable fraction of forecasts has naive expectations. In other words, participants anticipate that today's and future prices will mirror the price observed yesterday. Interestingly, the fraction of naive forecasts declines sharply following the shocks or their announcement, pointing to an increase in "eductive" reasoning. Second, participants revise their predictions when they observe a discrepancy between their past expectations and realized prices, but they do so sluggishly. For example, even if yesterday's price prediction was higher than the price that was realized, they do not excessively lower their price forecast. Finally, participants consider the past inflation rate in shaping their expectations only when the game they play exhibits strategic complementarity. In such cases, prices are more likely to follow a nonstation-

[4] The existing learning-to-forecast experiments, such as Heemeijer et al. (2009); Bao et al. (2017), demonstrate that, on one hand, prices deviate substantially from those expected under REE when there is a strong positive feedback between expectations and prices, and on the other hand, prices converge quickly to REE prices when there is a negative feedback between the two. Such an effect of strategic environment has also been demonstrated in a price-setting game (Fehr and Tyran, 2008; Funaki et al., 2023), a duopoly game (Potters and Suetens, 2009), and one-shot beauty contest games (Sutan and Willinger, 2009; Hanaki et al., 2019). Recently, however, Evans et al. (2022) and Anufriev et al. (2022a) demonstrate that eliciting forecasts for a longer horizon stabilizes the price dynamics under strategic complementarity, although not to the extent that prices converge to REE prices. While the frameworks of Evans et al. (2022) and Anufriev et al. (2022a) are different from ours, we would like to investigate whether the effect of a strategic environment is observed in our framework.

ary process or deviate from REE prices persistently. Therefore, it would be optimal for participants to incorporate past inflation rates in their predictions. In contrast, under strategic substitution, realized prices tend to be closer to the steady-state REE prices, motivating participants to base their predictions solely on the market price *level*s, not its change.

From these findings, we draw the following conclusions. There is not a single, universally correct model for expectation formation. Factors such as economic conditions, the presence of shocks, or their announcements seem to prompt participants to adaptively choose their predictive models. Similarly, participants are not purely backward-looking or forward-looking. Some participants only form forward-looking expectations when the shocks or their announcements occur. The studies most closely aligned with our paper's conclusions are those by authors Anufriev and Hommes (2012) and Evans et al. (2021). In these models, participants endogenously choose a prediction model that enhances the accuracy of their forecasts.

To the best of our knowledge, only a few, mostly recent, papers elicit forecasts for multiple future periods in the framework of learning-to-forecast experiments.[5] Colassante et al. (2020) elicit a series of forecasts for multiple future periods, however, unlike our framework, the forecasts beyond that for the next period do not determine the market outcome.

Evans et al. (2022) study the impact of forecasting horizon on the stability of the aggregate outcome. Their experiment is based on Lucas' asset pricing model (Lucas, 1978). Participants submit their forecast for *the average price over multiple future periods*, instead of the forecasts about the prices for each of these future pe-

[5] There are, however, several papers that elicit forecasts for multiple future periods in the asset market experiments pioneered by Smith et al. (1988). In these experiments, participants do not only forecast future prices, but actually trade the asset; see, e.g., Haruvy et al. (2007) and Akiyama et al. (2014, 2017).

riods, which determines the prices. They mix the short-horizon (those who forecast only one period ahead) and long-horizon (those who forecast over one to ten periods ahead) participants, and change the fraction of each across treatments. They find that while markets only with short-horizon forecasters exhibit substantial and pro-longed deviation from the REE, those markets with even a modest share of long-horizon forecasters converge.

In Adam (2007), Rholes and Petersen (2021), and Petersen and Rhoes (2022), they study a New Keynesian learning-to-forecast experiment. In their experiments, the one- and two-period-ahead forecasts of inflation rates elicited from participants determine the current outcome. These studies differ from our paper in how they model the effects of these forecasts on today's equilibrium outcomes. In Rholes and Petersen (2021); Petersen and Rhoes (2022), two forecasts impact the outcome in the opposite way. The one-period ahead forecast is positively related to the current outcome, the two-period ahead forecast is negatively related.

The paper most closely related to ours is Lustenhouwer and Salle (2022). Their experiment is based on a New Keynesian model with an inflation-targeting interest rate rule and a government sector. Participants submit their forecasts for output (in terms of its percentage deviation from the "normal" level) in multiple future periods. Like our study, these forecasts determine the current outcome. Furthermore, they study the impact of the announcement regarding future policy changes on the expectations as we do in our experiment. In this study, participants receive only qualitative information regarding how their forecasts, along with policy variables, collectively influence the output. In our study, however, participants do know how their forecasts collectively affect the outcomes. Also, we aim to complement this study by studying the strategic environment effects, which is not the focus of the paper.

The rest of the paper is organized as follows. Section 2 presents a model of a multi-period beauty contest based on Calvo (1983). The design of the experiment, as well as its procedure, are presented in Section 3. Section 4 shows the REE as the benchmark. The results of the experiment are summarized in Section 5, including a discussion of implications for modeling expectation formation. Section 6 concludes.

# 2 A Model of a Multi-Period Beauty Contest

We introduce our economic model, which serves as a basis for our experiment design. Suppose that there is a continuum of firms that have monopolistic power uniformly distributed over $[0, 1]$. Let $p_i$ denote the price chosen by individual firm $i$. For notational simplicity, we drop subscript $i$. The demand function for an individual firm is

$$D\left(p; P\right) \equiv \left[a - bp + cP\right]^{+}, \tag{1}$$

where $P$ is the aggregate price given by

$$P = \int_0^1 p_i di. \tag{2}$$

We assume that $a > 0$, $b > 0$, and $c \in \mathbb{R}$. We impose the following condition on $b$ and $c$ :

$$-\infty < c \leq 2b.$$

This assumption guarantees that the single-period version of this game has a REE price. Note that parameter $c$ can be positive or negative, which governs the degree of strategic interaction between the firms. The firms have an identical linear technology function, and the unit cost of production is denoted by $\kappa$.

Firm $i$ maximizes the present value of its profit in each period but is subject to pricing friction. The friction prevents the firms from changing its price in every period and only allows the firm to do so with probability $1-\theta$. We assume that if a firm cannot change its prices $T$ consecutive times, it can change the price next time with probability 1. Moreover, we assume that with a probability of $\gamma \in [0,1)$, all the firms are forced to exit from the market. This assumption is later used to mimic the indefinite-horizontal nature in our actual experiments conducted by participants, although this assumption plays no role in the theory.

The maximization problems in period $t$ for firms which can reset their prices are

$$\max_{p_t} \quad \sum_{s=0}^{T} ((1-\gamma)\theta)^s \pi_{t+s}, \tag{3}$$

where

$$\pi_{t+s} = (p_t - \kappa) \, D(p_t; P_{t+s}) \, .$$

The demand function, $D(p_t; P_{t+s})$, is given by equation (1). Recall that in every period, firm $i$ can re-optimize its price with probability $1-\theta$. Since the continuation payoffs in these events are not a function of $p_t$, they do not appear in objective function (3).

We establish the following proposition:

**Proposition 1.** *The optimal price for firms that can reset their prices at period $t$ is given by*

$$p_t = \sum_{s=0}^{T} \frac{((1-\gamma)\theta)^s}{\sum_{s=0}^{T}((1-\gamma)\theta)^s} \left(\alpha + \beta P_{t+s}\right), \tag{4}$$

*where*

$$\alpha = \frac{1}{2}\left(\kappa + \frac{a}{b}\right) \qquad \beta = \frac{1}{2}\frac{c}{b}.$$

*Moreover, the law of motion of the aggregate price (REE price) is*

$$P_t = (1-\theta)\, p_t + \theta P_{t-1}. \tag{5}$$

*Proof.* Taking the first-order conditions of maximization problem (3), we obtain equation (4). To derive equation (5), recall that the aggregate price is given by the average of individual prices (2) and that only fraction $1 - \theta$ of firms can reset their prices. Because they are randomly chosen, the average price among firms that cannot reset their price today is $P_{t-1}$. Firms that can reset their prices choose the same price level given by equation (4). So, average price $P_t$ satisfies equation (5). □

It is important to note that $\alpha + \beta P_t$ corresponds to the optimal flexible price in period $t$, if the firms can reset their prices. Therefore, the optimal price (4) constitutes a weighted average of the optimal flexible prices. Equation (4) elucidates the dynamic thought process within firms: the optimal price, $p_t$, depends on current actions by others, as well as future actions, encapsulated in $\mathbf{P} = (P_{t+s})_{s=0}^{T}$. In other words, the best response function, given by (4), embodies dynamic feedback mechanisms. The aggregate price today is influenced by future prices. This is the primary reason why this model is frequently described as a *dynamic beauty contest model* (e.g., Angeletos and Lian, 2018).

In contrast, when firms can fully adjust their prices ($\theta = 0$ or $T = 0$), the dynamic component of the decision-making process is eliminated: each firm's focus shifts solely to predicting the present-day actions of the other firms. The experiment we propose, inspired by this economic model, stands apart from the existing single-period one because it necessitates that participants anticipate not only the actions of others in period $t$, but also in later periods $t + 1, \cdots, t + T$.

## 3 Experimental Design

### 3.1 Setup

In our experimental setup, we employ a cohort of six participants.[6] These participants indexed by $i$ engage in the game over a series of periods. As each period begins, the participants put forward their price forecasts for five subsequent periods, including the current one.[7] For example, in the first period, participants submit their price forecasts for periods 1 through 5. In the second period, forecasts are given for periods 2 through 6, and this pattern continues for subsequent periods. We denote each participant's forecast submitted in period '$t$' for the price in period '$k$' as $f^i_{t,k}$. Importantly, not all submitted forecasts affect the participants' rewards, as will be explained below. However, at the beginning of a period, when participants submit their forecasts, all five of these forecasts potentially influence their rewards.

Let $\pi^i_t$ denote the payoff of participant $i$ at period $t$. It is determined by

$$\pi^i_t = \frac{100}{|F^i_t - P_t| + 1},$$

where $P_t$ is the realized price in period $t$ and $F^i_t$ denotes the payoff-relevant forecast of participant $i$ at period $t$, which is recursively defined.[8] In the first period $t$, the payoff-relevant forecast $F^i_t$ is set to $F^i_1 = f^i_{1,1}$. In period 2, with probability $1 - \theta = 1/2$, the newly submitted set of forecasts in this period is used for determining $F^i_2 : F^i_2 = f^i_{2,2}$. With probability $\theta = 1/2$, the forecasts submitted in period 1 is used for determining $F^i_2$. In this case, $F^i_2 = f^i_{1,2}$. The payoff-relevant sets of forecasts for $t = 3$ and onward are determined in the same manner. When a set of forecasts has been payoff-

[6] One may consider a cohort of six to be too small for an experiment to have macroeconomic implications. However, the main results of existing learning-to-forecast experiments do not change even if conducted with larger groups of 20 to 30 participants (Bao et al., 2020) or with close to 100 participants (Hommes et al., 2021).

[7] All the forecasts are constrained to be integers. We provide a rationale for our choice of the horizon, $T = 4$, in footnote 12 below.

[8] $\pi^i_t$ is rounded to the nearest integer value. This way of incentivizing the forecast accuracy is also used, for example, in Adam (2007); Assenza et al. (2021); Anufriev et al. (2022b).

Table 1: Hypothetical Submitted Forecasts

| Period\Forecast Period | $t = 1$ | $t = 2$ | $t = 3$ | $t = 4$ | $t = 5$ | $t = 6$ | $t = 7$ | $t = 8$ | $F_t^i$ |
|---|---|---|---|---|---|---|---|---|---|
| $t = 1$ | **10** | **11** | 12 | 12 | 12 | | | | **10** |
| $t = 2$ | | ~~12~~ | ~~11~~ | ~~12~~ | ~~13~~ | ~~13~~ | | | **11** |
| $t = 3$ | | | **9** | 10 | 11 | 10 | 10 | | **9** |
| $t = 4$ | | | | **13** | 12 | 12 | 10 | 10 | **13** |

relevant for five consecutive periods, then with probability one, the new set of forecasts submitted in the next period becomes payoff-relevant.[9]

This adjustment process is motivated by the multi-period beauty contest model in Section 2. The probability of firms being given an opportunity to re-optimize their prices in period $t$, $1 - \theta$, is translated into the probability of the new set of forecasts becoming payoff-relevant for participants. And the horizon over which firms optimize, $T$, in the model is equivalent to the number of future periods, in addition to the current one, over which our participants forecast in each period. In the case of our experiment, $T = 4$. We set the payoffs so that the equilibrium path of the Nash equilibrium (subgame perfect Nash equilibrium) corresponds to the REE of the multi-period beauty contest game in Section 2.

To give a concrete example, consider hypothetical sets of forecasts in Table 1. Each entry represents a submitted forecast. For example, in period $t = 1$, the participant submits her forecasts of the prices of period $t = 1$ to 5 as follows:

$$\left(f_{1,1}^i, f_{1,2}^i, f_{1,3}^i, f_{1,4}^i, f_{1,5}^i\right) = (10, 11, 12, 12, 12)\,.$$

In $t = 1$, this set of forecasts becomes payoff-relevant: $F_1^i = 10$. Suppose that in

[9]This is because, in such cases, the set of forecasts that has been payoff-relevant does not contain the forecast for the next period. For example, if the set of forecasts submitted in period 1 $\left(f_{1,1}^i, f_{1,2}^i, f_{1,3}^i, f_{1,4}^i, f_{1,5}^i\right)$ has been payoff-relevant for five consecutive periods, i.e., between periods 1 to 5, then as it does not contain a forecast for period 6 price, we need to make the new set of forecasts submitted in period 6 payoff-relevant.

period $t = 2$, the new set of forecasts does not become payoff-relevant. So, $F_2^i$ is determined based on the forecasts submitted in period $t = 1$: $F_2^i = 11$. In periods $t = 3$ and $t = 4$, as the new set of forecasts becomes payoff-relevant, $F_3^i = 9$ and $F_4^i = 13$.[10]

Now we move on to explain how the aggregate price at period $t$, $P_t$, is determined. Motivated by the model in Section 2, the aggregate price is given by

$$P_t = \frac{1}{6}\left( \sum_{i \text{ cannot reset}} p_{t-1}^i + \sum_{i \text{ can reset}} \sum_{j=0}^{4} \frac{(1-\gamma\theta)^j}{\sum_{l=0}^{4}(1-\gamma\theta)^l} \left(\alpha + \beta f_{t,t+j}^i\right) \right), \tag{6}$$

where $\alpha$ and $\beta$ are the parameters of the model, $\theta = 1/2$, and $\gamma = 0.05$.[11] Here, equation (6) is the empirical-counterpart of equation (2). Some of the participants cannot change their prices, which are captured by the first term of the RHS in the above equation. Moreover, the mapping between the submitted forecasts to the optimal choice, the second term, is motivated by equation (4).[12]

On the screen in which participants submit their forecasts, the values of $\alpha$ and $\beta$ are presented clearly. On the same screen, participants are informed of the realized $P_t$ as well as the payoff-relevant forecast $F_t$ in all the past periods. See Appendix C for the screenshots.

In our experiment, we assume that the game ends with probability $\gamma = 0.05$ at the

[10]If the new sets of forecasts in period 3 and 4 do not become payoff-relevant, then $F_3^i = f_{1,3}^i = 12$ and $F_4^i = f_{1,4}^i = 12$.

[11]In the experiment, however, due to an oversight, the price was determined with $\gamma = 0$ instead of $\gamma = 0.05$. The price determination equation with the wrong value of the weights was also communicated to participants in the instruction. As we show in Appendix A, however, this oversight would result in the REE price to differ only slightly (by less than 1). Furthermore, because $P_t$ is rounded to the nearest integer in our experiment as the forecasts submitted by participants are constrained to be integers, we believe this oversight would not change the experimental outcomes in any major way.

[12]Equation (6) provides us a rationale for our choice for $T = 4$. Note that when $\theta = 1/2$ and $\gamma = 0.05$, the impact of $f_{t,t+4}^i$ on the optimal price is minimal. This is because the weight for $f_{t,t+4}^i$ is $((1-\gamma)\theta)^4 / \sum_{k=0}^{4}((1-\gamma)\theta)^k$ , which is approximately 2.7%. Therefore, allowing participants to make longer forecasts is unlikely to change the results. This argument does not hold if we choose a lower adjustment probability, $\theta$.

end of each period. Otherwise, the game continues to the next period. Participants are rewarded based on the sum of points they earn from the beginning until the end of the game.

However, implementing such a probabilistic termination rule in a laboratory experiment is problematic because the game can end after only a few periods, thereby not allowing us to study the evolution of the forecasts in response to shocks, or the game may not end during the time for which participants are recruited. To circumvent the former problem, we employ the block random termination method Fréchette and Yuksel (2017) commonly used in the experiments of indefinitely repeated games.

Under this method, participants are asked to play the game in a block of $B$ periods. Only at the end of the block are the participants informed of whether the game has already ended during the block just completed.[13] If the game has done so, they gain the sum of $\pi_t^i$ up to that period. If the game has not ended, the game continues for another block of $B$ periods. For example, if the game has actually ended in period $\tau < B$, they are rewarded based on $\sum_{t=1}^{\tau} \pi_t^i$. They are also well informed, the game can continue beyond B periods, and if that is the case, they play the game for at least another B periods. In our experiment, we set $B = 20$. Therefore, participants are informed that they are going to play the game for at least 20 periods.

Furthermore, because the duration of a game may vary across groups, participants are informed that this multi-period game can be repeated several times if all the groups finish playing a game within 30 minutes from the start of the first game. When a new game starts, new groups of six participants are randomly formed (and thus, everyone in the same experimental session has to wait until all the groups in the same session finish a game before starting a new one), and they play the game with the

[13] In our experiment, this information is communicated to the participants with the sequence of random numbers that determine the termination of the game in each period displayed on the same screen.

same parameter values.[14] If the game is repeated several times, one of them will be selected randomly, and participants are paid according to the points they earn in the chosen game.

For example, if the last group to finish the first game does so within 15 minutes from its start, the second game will be played. If the last group to finish the second game does so at 35 minutes after the start of the first game, the experiment ends. In this case, because the game has been repeated twice, either the first or the second game is selected randomly, and the participants are paid based on their earnings in the selected game. Note that it is also possible that the first game ends 30 minutes after its start. If this is the case, only one game is played during the experiment, and the participants are paid according to the points they earned in this game.

During the games, we introduce (a maximum of) two shocks to $\alpha$, which captures the size of the demand during the first two blocks of 20 periods (one shock in each block). The literature suggests, even under strategic substitution where prices often converge to the steady state level, it may take several periods to do so. To allow for the prices to stabilize before introducing a shock, we have introduced the first shock at the beginning of period 14, and the second shock, in case it is reached, at period 29. We choose $\alpha$ so that the initial flexible REE price is 65, and it becomes 85 and 110 after the first and the second shocks, respectively.[15]

[14] Because the game was repeated with the same parameter values, we have opted for random re-matching to avoid participants coordinating on the realized prices in the earlier games.

[15] One could consider introducing shocks differently, for example, to introduce a positive shock that is followed by a negative shock. We have opted for two positive shocks to investigate whether participants better adjust to the second shock, if they reach it, as they have gained experience through the first shock.

## 3.2 Treatments

In our 3 by 2 between-participants experiment, we focus on modulating two primary aspects of the games. The first aspect is the degree of strategic interaction. We consider the games with $\beta \in \{0.9, -0.9, -1.8\}$ . That is, the games exhibit strategic complementarity (positive feedback) or substitution (negative feedback).[16] We also explicitly consider a strong substitution case, $\beta = -1.8$. This is because New-Keyenesian models might exhibit such strong substitutability. García-Schmidt and Woodford (2019) propose an expectation formation models allow this kind of strong substitutability.

The second aspect is the anticipation of the shocks to $\alpha$. We consider treatments with and without pre-announcement of shocks. In the treatment without pre-announcement, participants are informed of the new value of $\alpha$ only at the occurrence of the shock, that is, in period 14 for the first shock and in period 29 for the second shock. In the treatments with pre-announcement, participants are informed of the new value of $\alpha$ at period 12 for the first shock, and at period 27 for the second shock. Thus, we can study the adjustment of participants' forecasts, say for period 14 price, before and after the pre-announcement.

## 3.3 Procedures

We use oTree (Chen et al., 2016), an open-source platform for web-based interactive tasks, to conduct the experiments online. Participation is from where they are, and not from our physical experimental laboratory. We used Zoom to manage the

[16] Because participants are only allowed to submit integer forecasts, there might be multiple equilibria for cases of strong strategic complementarity. It turns out that when $\beta = 0.9$, there are multiple equilibria. However, the set of equilibria is not huge and this multiplicity does not affect the results of the paper. Therefore, we do not address this issue explicitly.

experiments.[17]

Once ready and after a general instruction about the online experiment is given, the prerecorded instruction video is played on the screen. Although participants are not given a hard copy of the instruction slides, they are informed that they can go through the same set of slides after the video finishes until they finish answering the comprehension quiz. All the participants need to answer all six questions of the quiz correctly for the first game to start. As noted, participants can review the instruction slides before and while answering the quiz. The final rewards are paid through Amazon Gift Cards (e-mail version).

Participants are students at Osaka University. They are recruited using ORSEE (Greiner, 2015). An English translation of the instruction slides, examples of the decision screens, and the comprehension quiz are provided in Appendix C.

## 4 Rational Expectations Hypotheses

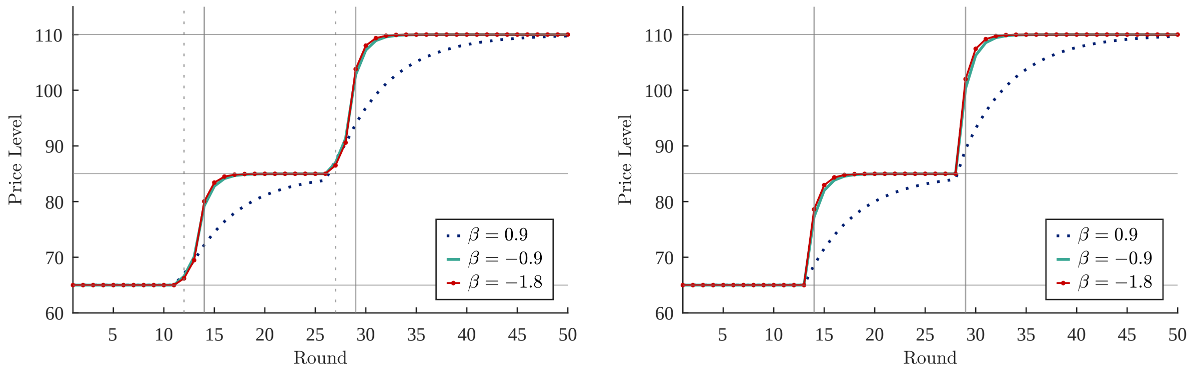
We take rational expectations as our benchmark. Given the initial conditions, $P_{-1}$, REE prices $\left(p_t^{REE}, P_t^{REE}\right)$ solve the following equations uniquely:

$$P_t^{REE} = (1-\theta)\, p_t^{REE} + \theta P_{t-1}^{REE} \tag{7}$$

$$p_t^{REE} = \sum_{k=t}^{T-t-1} \frac{((1-\gamma)\theta)^{k-t}}{\sum_{s=0}^{T}((1-\gamma)\theta)^s} \left(\alpha_{t,k} + \beta P_{t+k}^{REE}\right), \tag{8}$$

where $\alpha_{t,k}$ is the expected value of $\alpha$ at period $k$ from a period $t$ perspective. When the shocks are not announced, $\alpha_{t,k} = \alpha_{t,t}$ for all $k$ and $t$. Equation (7) describes the law of motion of $P_t^{REE}$ and $p_t^{REE}$ is the optimal price subject to the pricing friction. Figure 1a demonstrates the REE price sequence $\left(P_t^{REE}\right)_t$ when the shocks are pre-announced, and Figure 1b depicts the price sequence when the shocks are not

[17] See Appendix B for details regarding how we have conducted our online experiment.




(a) With Announcement (b) Without Announcement

Figure 1: Price Dynamics under Rational Expectations

pre-announced.

These figures are intuitively understood. When the game demonstrates strategic complementarity ($\beta > 0$), the transition to new steady-state equilibrium prices is slower, a feature resulting from Calvo pricing friction. Participants know that others might not be swiftly adjusting their prices due to the pricing friction. As individual optimal prices are positively correlated with the pricing actions of others, they prefer to adjust their prices slowly.

On the contrary, when the game displays strategic substitution ($\beta < 0$), the operative mechanism is reversed. Given that some participants fail to adjust their prices, the prevailing price becomes too low. This low price motivates individuals to post higher prices. Thus, the transition to the new steady-state equilibrium price is faster.

Table 2: Number of Groups per Treatment

| Treatments | $\beta = 0.9$ | $\beta = -0.9$ | $\beta = -1.8$ |
|---|---|---|---|
| With Announcement | 8 | $8^a$ | 8 |
| Without Announcement | 8 | 8 | $7^b$ |

*Notes*: Row "With Announcement" represents the results where the shocks were pre-announced while row "Without Announcement" shows the results where the shocks were not pre-announced. $a$: One group stopped at $t = 33$ due to a technical problem. $b$: One group stopped at t=39 due to a technical problem.

# 5 Results of the Experiment

We conducted our experiments in April and May 2023. There was a total of 294 participants in the experiments.[18] Table 2 summarizes the number of groups per treatment. Each experiment lasted for 90 minutes on average, and participants earned 2482 JPY (approximately 18 USD based on the exchange rate at the time) including the show-up fee of 500 JPY on average.[19] The average payment varied across the value of $\beta$. It was lowest in the treatments with $\beta = -1.8$ (1778 JPY) followed by 2806 JPY and 2844 JPY in treatments with $\beta = -0.9$ and $\beta = 0.9$, respectively.

[18]However, in one session with announcement of the future shock with $\beta = 0.9$, one participant decided to leave the experiment while answering the comprehension quiz. The experimenter has substituted this participant, and thus, the experiment could continue. We have dropped the data of this group from the analyses. But because we did the same treatment in which 30 participants showed up, we have a total of eight groups for this treatment. Furthermore, in one session without announcement of the shock and $\beta = -1.8$, one of the participants lost his/her internet connection around period 11 and switched to a different mode of connection. As a result, an error occurred and the experiment could not continue for this group. In addition, in the same session, an error occurred for a group in period 40. Thus, we only have data for seven groups for this treatment, with one group missing the price and forecasts submitted in period 40. Finally, in the treatment with announcement of the shocks and $\beta = -0.9$, one group encountered a technical problem and the experiment stopped in period 33. Thus, while we have data for eight groups for this treatment, one lacks the data from period 34 onward.

[19]Exchange rate between the points earned during the experiment and JPY was 1 point = 2 JPY.

## 5.1 Effect of the Degree of Strategic Interaction

Figure 2 shows the dynamics of prices observed in each treatment. One line corresponds to a group in each panel. As one can observe, regardless of the existence of the announcement, the prices follow the REE prices quite closely when $\beta = -0.9$. When $\beta = -1.8$, the prices are close to the flexible REE prices but fluctuate greatly around them. These fluctuations are also in sharp contrast to the rational expectations outcome presented in Figure 1. When the game exhibits strategic complementarity, $\beta = 0.9$, they persistently deviate from the REE prices. These findings are very similar to the results from the existing learning-to-forecast experiment such as those of Heemeijer et al. (2009) and Bao et al. (2012). The equilibrium outcomes are more stable if the game exhibits strategic weak substitutability, $\beta \in (-1, 0]$.

Now, let us quantify the degree of deviation from the REE prices. We compute the relative absolute deviation, $RAD$, and the relative deviation, $RD$, proposed by Stöckl et al. (2010). For group $g$, $RAD_g$ and $RD_g$ are calculated as follows:

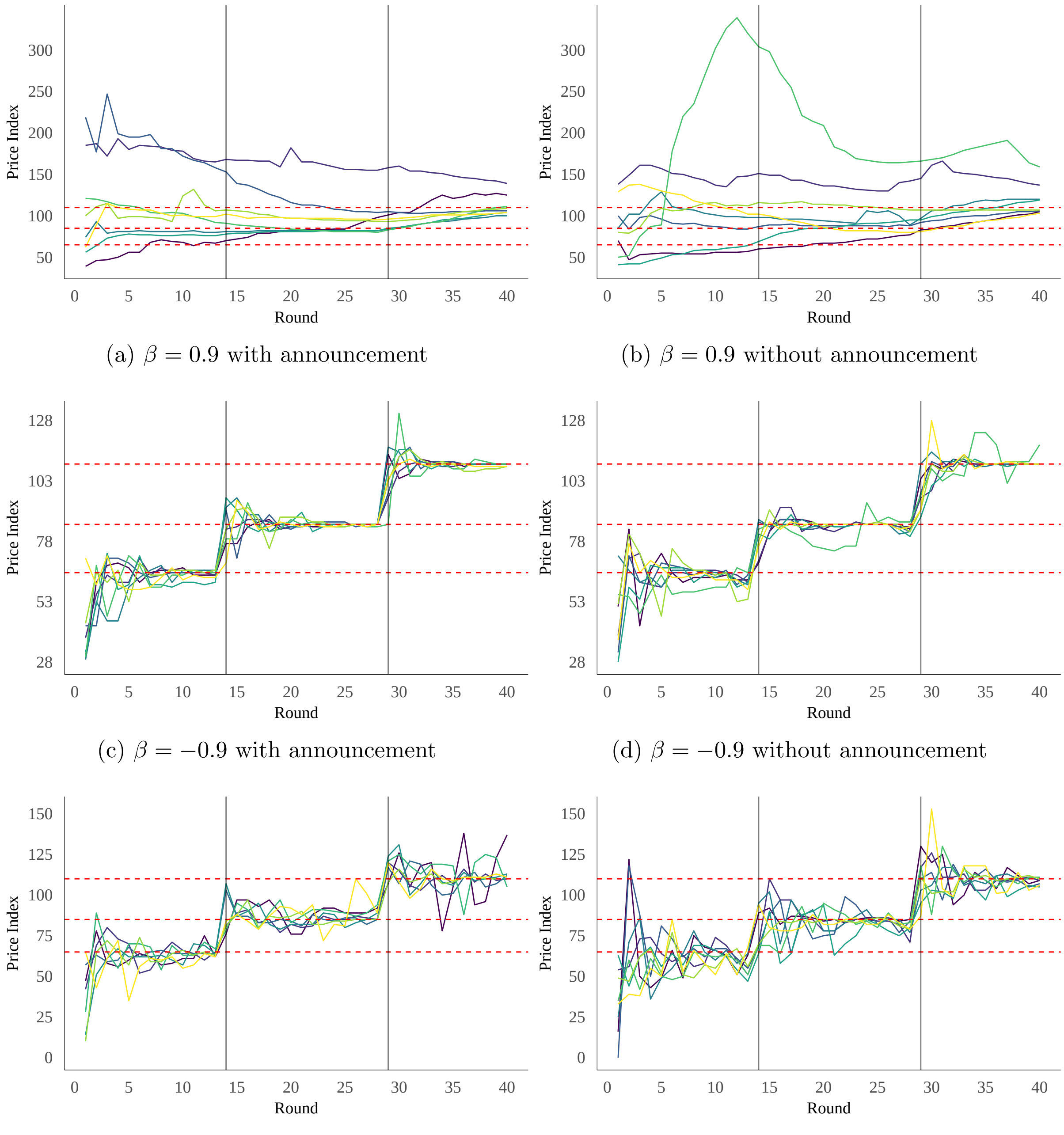
$$RAD_g = \frac{1}{K} \sum_t \frac{\left| P_{g,t} - P_t^{REE} \right|}{P_t^{REE}}$$

$$RD_g = \frac{1}{K} \sum_t \frac{P_{g,t} - P_t^{REE}}{P_t^{REE}}$$

where $P_{g,t}$ is the realized period $t$ price for group $g$ and $P_t^{\text{REE}}$ is the REE price in period $t$.[20] $K$ is the total number of periods (40 except for the two groups that faced a technical problem).

Figure 3 shows the empirical cumulative distribution (ECD) of RAD (top) and RD (bottom) in the treatments with (left) and without (right) announcement. In each panel, the distributions for each $\beta$ are shown. The top panels show that, regard-

[20] Here we use the REE price based on the price determination equation used in the experiment, namely, the one that set $\gamma = 0$. See, footnote 11.




(a) $\beta = 0.9$ with announcement

(b) $\beta = 0.9$ without announcement

(c) $\beta = -0.9$ with announcement

(d) $\beta = -0.9$ without announcement

(e) $\beta = -1.8$ with announcement

(f) $\beta = -1.8$ without announcement

Figure 2: Realized Price Dynamics

*Notes*: The red dashed lines represent the steady-state levels of the aggregated prices.

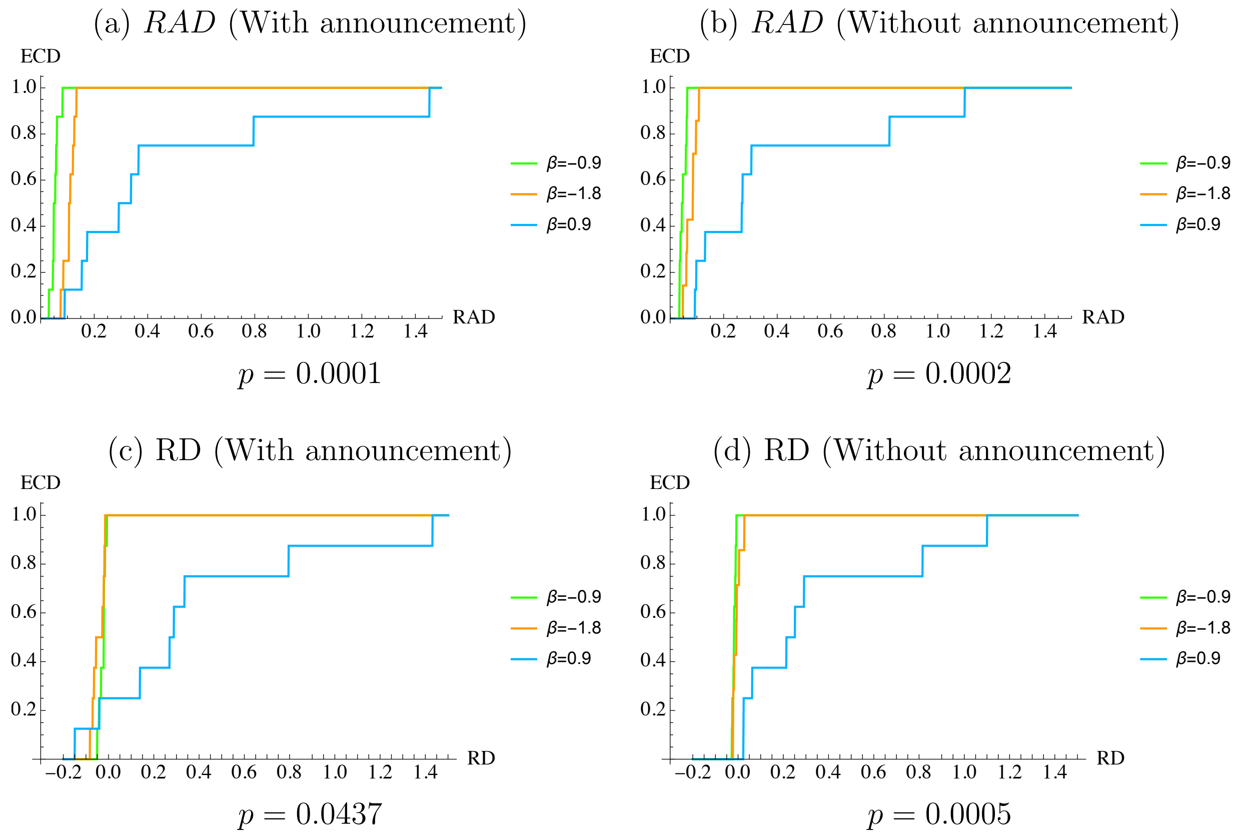



Figure 3: Distribution of RAD (top) and RD (bottom)

*Note*: P-values of Kruskal–Wallis (KW) test are reported.

less of the existence of announcement, RADs are positive for each $\beta$. Applying the single-rank test, the observed prices are significantly different from the REE prices.[21] Furthermore, the observed RADs are largest under $\beta = 0.9$, followed by $\beta = -1.8$, and then $\beta = -0.9$ regardless of the existence of the announcement. The differences between the three treatments are statistically significant both with and without announcement.[22] Therefore, the deviations of the observed prices from the REE

[21] They are all significantly different from zero at the 5% level according to the signed-rank test. $P - values$ are 0.018 $\beta = -1.8$ without announcement and 0.012 for all the other five treatments.

[22] $P < 0.05$ based on the Kruskal–Wallis (KW) test for both cases. For pairwise comparisons, p-values based on the Mann–Whitney (MW) test (two-tailed) are 0.0002 ($\beta = 0.9$ vs $\beta = -0.9$), 0.0022 ($\beta = 0.9$ vs $\beta = -1.8$), and 0.0093 ($\beta = -0.9$ vs $\beta = -1.8$) without announcement, and 0.0002 ($\beta = 0.9$ vs $\beta = -0.9$), 0.0047 ($\beta = 0.9$ vs $\beta = -1.8$), and 0.0003 ($\beta = -0.9$ vs $\beta = -1.8$) with announcement.

prices are smaller under strategic substitution than under strategic complementarity, and furthermore, they are smaller under "weak" substitution ($\beta = -0.9$) than under "strong" substitution ($\beta = -1.8$).

The bottom panels show, however, that while RDs are larger under strategic complementarity than under strategic substitution as in the case of RADs, there is no longer a statistically significant difference between $\beta = -0.9$ and $\beta = -1.8$.[23] Note that because RD takes the direction of the deviation from REE into account, it becomes close to zero when the prices fluctuate around REE as in the case of $\beta = -1.8$.[24]

## 5.2 Forward-Lookingness of Expectations

Let us now turn to the way forecasts have responded to the announcement of future shocks. To do so, we focus on the difference in the forecast adjustments at the time of shock between treatments with and without announcement. Namely, we compute the average difference in the forecast price between the period of the shock and the period just before the shock (periods 14 and 13, respectively, for the first shock and periods 29 and 28, respectively, for the second shock) submitted in the periods after the announcement (periods 12 and 13 for the first shock and periods 27 and 28 for the second shock) and those before the announcement (periods 10 and 11 for the first shock and periods 25 and 26 for the second shock). That is, for each participant $i$, the average forecast responses to the announcement of the first and second shocks,

[23] For pairwise comparisons, p-values based on the MW test (two-tailed) are 0.0002 ($\beta = 0.9$ vs $\beta = -0.9$), 0.0012 ($\beta = 0.9$ vs $\beta = -1.8$), and 0.1520 ($\beta = -0.9$ vs $\beta = -1.8$) without announcement, and 0.0650 ($\beta = 0.9$ vs $\beta = -0.9$), 0.0379 ($\beta = 0.9$ vs $\beta = -1.8$), and 0.2345 ($\beta = -0.9$ vs $\beta = -1.8$) with announcement.

[24] In fact, RD is not significantly different from zero for $\beta = -1.8$ without announcement (p=0.6875, signed-rank test) while it is significantly different from zero at the 5% significance level for five other treatments (p=0.0391 for $\beta = 0.9$ with announcement and 0.0078 for the remaining four treatments, signed-rank test.)

$\Delta f_1^i$ and $\Delta f_2^i$, are defined as follows:

$$\Delta f_1^i = \frac{1}{2}\sum_{t=12}^{13}\left(f_{t,14}^i - f_{t,13}^i\right) - \frac{1}{2}\sum_{t=10}^{11}\left(f_{t,14}^i - f_{t,13}^i\right)$$
$$\Delta f_2^i = \frac{1}{2}\sum_{t=27}^{28}\left(f_{t,29}^i - f_{t,28}^i\right) - \frac{1}{2}\sum_{t=25}^{26}\left(f_{t,29}^i - f_{t,28}^i\right).$$

We compute the same measure for the participants in the treatment without announcement as for the benchmark.

Figure 4 shows the empirical cumulative distributions of $(\Delta f_1^i)_i$ (left) and $(\Delta f_2^i)_i$ (right) with and without announcement for three values of $\beta$. P-values are from MW and Kolmogorov–Smirnov (KS) tests. We also report the distributions between treatments with and without announcement.

Figure 4 shows that, without announcement, both $\Delta f_1^i$ and $\Delta f_2^i$ are concentrated around zero. This is especially so for $\beta = -0.9$ for both the first and second shocks in which prices were very close to the steady-state equilibrium prices before the realization of these shocks. With announcement, however, $\Delta f_1^i$ and $\Delta f_2^i$ are distributed more toward positive values (except for $\Delta f_2^i$ of $\beta = 0.9$), suggesting that participants adjusted their forecasts upward in response to the announcement. Except for $\beta = 0.9$, there are significant differences in $\Delta f_1^i$ and $\Delta f_2^i$ between treatments with and without announcement. Thus, consistent with participants having forward-looking expectations, they expect prices to go up after the announcement under strategic substitution.[25]

[25]While $\Delta f_2^i$ is significantly greater than $\Delta f_1^i$ for $\beta = -0.9$ ($p^{MW} = 0.041$ and $p^{KS} = 0.034$), there is no significant difference between the two for $\beta = -1.8$ ($p^{MW} = 0.968$ and $p^{KS} = 0.760$) and $\beta = 0.9$ ($p^{MW} = 0.311$ and $p^{KS} = 0.531$).

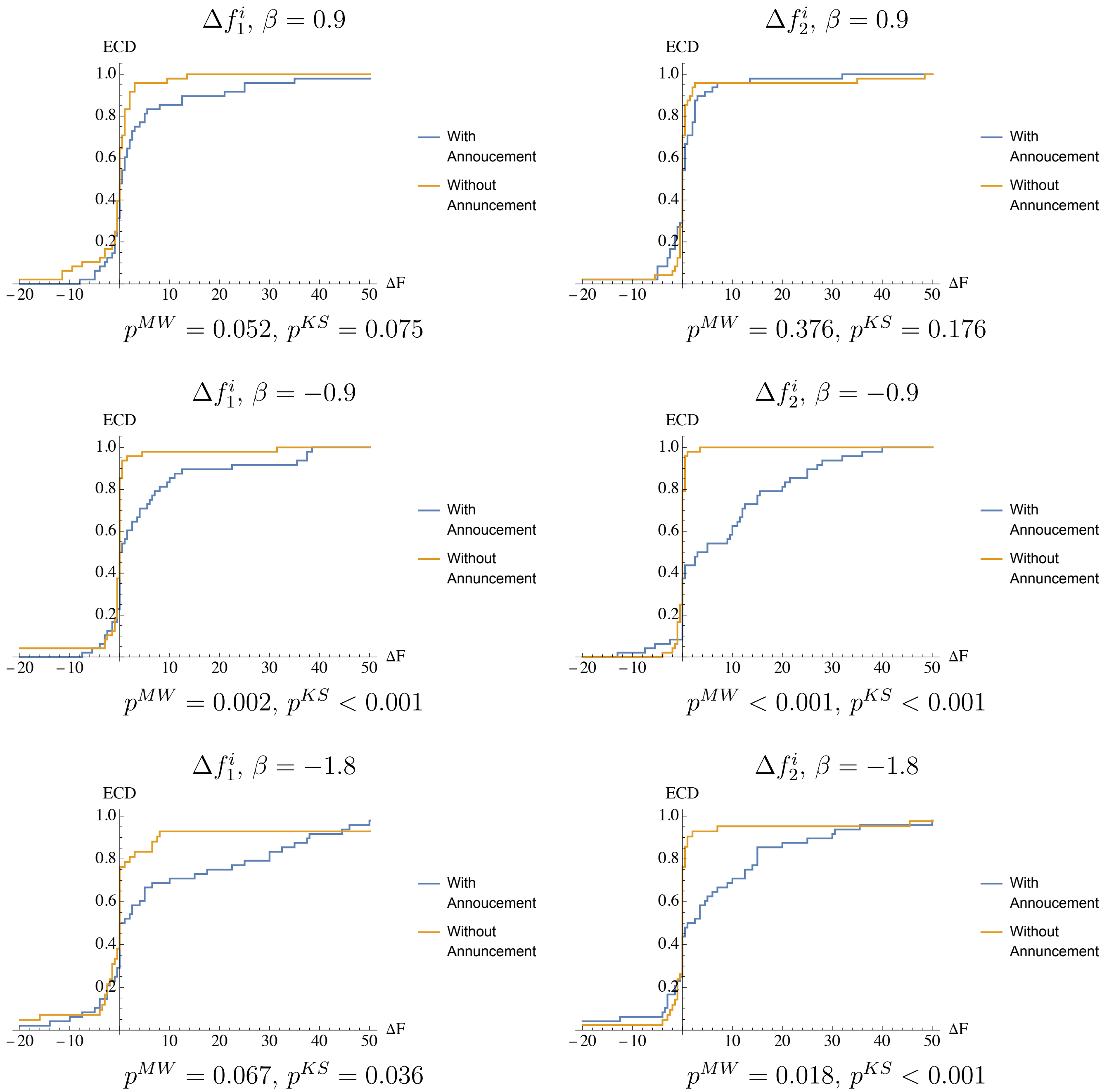



Figure 4: Empirical Distributions of $(\Delta f_1^i)_i$ and $(\Delta f_2^i)_i$

## 5.3 Expectation Formation in a Multi-Period Environment

In this section, we analyze how participants form expectations in our multi-period experiments. Similar to the existing literature, we categorize price expectations into two types: naive expectations and others. In Subsection 5.3.1, we define and analyze naive expectation formation in our setting. Following that, in Subsection 5.3.2, we proceed to analyze non-naive forecasts.

### 5.3.1 Naive Expectation Formation

We commence by defining a naive expectation. These definitions extend the notion of naive expectation from single-period environments to multi-period environments. We say that the forecasts of participant $i$ in period $t$, $\left(f^i_{t,t+j}\right)_{j\in\mathcal{J}}$, is *naive* if for all $j \in \mathcal{J}$,

$$f^i_{t,t+j} = P_{t-1},$$

and $\varepsilon$-*naive* if for all $j \in \mathcal{J}$,

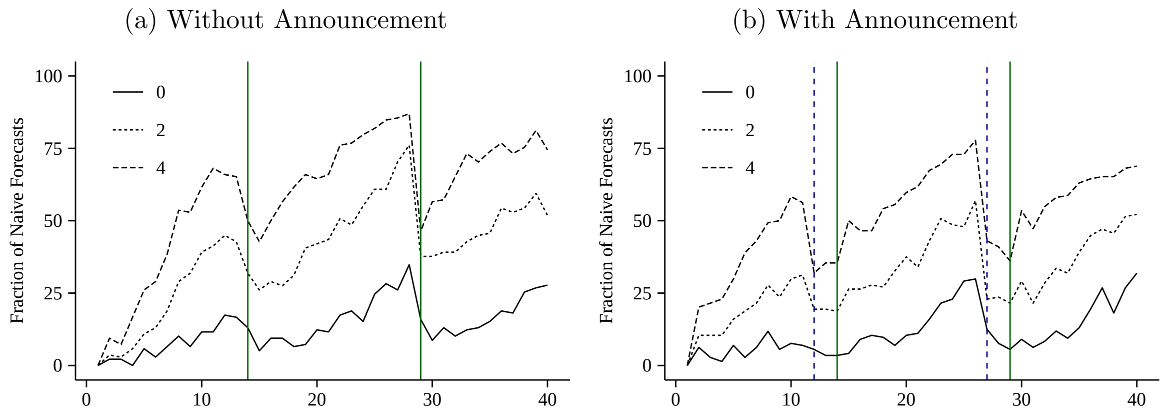
$$\left|f^i_{t,t+j} - P_{t-1}\right| \leq \varepsilon.$$

Figure 5 shows the fraction of the naive forecasts for $\varepsilon \in \{0, 2, 4\}$ in the treatment without the pre-announcements of shocks (left) and with the pre-announcements (right).[26] In each panel, the data from three values of $\beta$ are pooled. There are three points to be made about which kind of expectation formation participants will engage in. First, the number of participants forming naive expectations increases over time.[27] This is because there is a tendency for the economy to converge to steady-state REE prices. As a result, a growing number of participants form naive

[26]If we assume that prices had converged to a steady state, those exhibiting a deviation of approximately less than 5% from the steady state prices would be classified as naive when $\varepsilon = 4$.

[27]Under the REE hypothesis, the fraction of naive forecasts is zero after the realization of the shocks, and becomes 1 when the price sufficiently converges to the steady-state REE prices.

Figure 5: Fraction of Naive Forecasts for Various Cutoffs




*Notes*: The solid vertical lines represent the rounds at which the shocks occurred. The dashed vertical lines in the right figure represent the rounds at which the announcements of the shocks were made. The difference in line types represents the difference in the value of $\varepsilon$. As indicated in the legend, a solid line represents $\varepsilon = 0$, a dotted line indicates $\varepsilon$ being 2, and a dashed line 4.

expectations following the shocks or announcements endogenously.

Second, despite the shocks or announcements, a significant proportion of participants remain inclined toward naive expectations. To see this point, consider the fraction of naive forecasts when $\varepsilon = 2$. 25 percent of the forecasts are naive even right after the shocks or their announcement (Figures 5a and 5b).

Finally, a major chunk of participants either do not stick to, or entirely abandon naive expectation formation when faced with shocks or announcements. Instead, their expectation pattern leans more toward "eductive" tendencies, which will be elaborated on in the next section.

#### 5.3.2 Non-Naive Expectation Formation

We now pivot to non-naive expectations, basing our analysis on experimental data that is not $\varepsilon$-naive with $\varepsilon = 4$. While it is hard to identify how exactly each participant forms her forecast (e.g., level-1) due to the multi-period nature, we can still understand how participants, on average, form their forecasts. Drawing inspiration from Anufriev and Hommes (2012), we structure the forecast formation around two forecasting rules: adaptive heuristic and trend-following. If participant $i$ adopts the adaptive rule, their belief adjustment would proceed as follows:

$$f_{t,t+j}^{i} = \omega f_{t-1,t-1}^{i} + (1-\omega) P_{t-1} = \omega \left( f_{t-1,t-1}^{i} - P_{t-1} \right) + P_{t-1}, \tag{9}$$

where $\omega$ is the weight assigned to the forecast at period $t-1$. When $\omega \in (0,1)$, this forecasting rule says that they only adjust their forecast gradually. When $\omega < 0$, participant $i$ adjusts their forecast downward if the forecast at period $t-1$ is higher than the realized price at period $t-1$.

The trend-following rule says that participant $i$ assumes that the aggregate trend will persist:

$$f_{t,t+j}^{i} = P_{t-1} + \chi \left( P_{t-1} - P_{t-2} \right). \tag{10}$$

Parameter $\chi$ governs the strength of the extrapolation effect. When $\chi$ is negative, then we can interpret equation (10) differently. Namely, participant $i$ expects the short-run trend to be reversed.

We nest models (9) and (10) as follows and obtain the baseline regression specification:

$$f_{t,t+j}^{i} - P_{t-1} = a\Delta\alpha_t + 1_{A_t} + \gamma \underbrace{\omega \left( f_{t-1,t-1}^{i} - P_{t-1} \right)}_{\text{Adaptive Forecasting}} + (1-\gamma) \underbrace{\chi \left( P_{t-1} - P_{t-2} \right)}_{\text{Trend-Following Forecasting}} + \mathrm{FE}_i + \varepsilon_{t,i,j},$$

where parameter $\gamma$ assigns the weight to the adaptive forecasting rule, $\Delta\alpha_t$ is the shock, and $1_{A_t}$ is the indicator function that is one if at period $t$, the announcement to a shock is made. When $\gamma = 1$, then participant $i$ only uses the adaptive forecasting rule (9). If $\gamma = 0$, then they only use the extrapolation. By running the regression, we estimate the parameters $\gamma\omega$ and $(1-\gamma)\chi$. The multiplicative $\gamma\omega$ reflects the effective "weight" assigned to the adaptive forecasting rule.

The regression results for each forecast horizon $j$ are presented in Table 3. From these regression results, four points regarding expectation formation can be discerned. First, in almost all cases, participants' expectations are influenced by the shocks. Importantly, this reaction is not just limited to expectations at the time of the shock, $f^i_{t,t}$, but also for future expectations, $f^i_{t,t+j}$. Interestingly, in the case where $\beta = 0.9$, the shocks have a greater impact on future time points than on the current time point. Such predictions also hold when participants have rational expectations. On the other hand, no such empirical patterns are observed when $\beta < 0$. This implies that participants do not necessarily have qualitatively correct dynamic price forecasts if $\beta < 0$.

Second, in many cases, participants' predictions at the time of the announcement do not significantly respond to the announcement itself. This observation is evident from the fact that the coefficients for $1_{A_t}$ when $j = 0$ are insignificant. However, starting from $j = 2$, there are cases where the coefficient becomes significant. Participants anticipate that other participants will not immediately respond to the announcement itself, but they anticipate that others will react when the shocks occur two periods later. This finding is consistent with that in Section 5.

Third, it is evident that participants are employing adaptive-heuristic forecasting. Across all cases, it can be interpreted that when making price predictions, participants refer to the discrepancy between their past predictions and actual prices. Notably, the coefficients do not increase as the value of $j$ becomes larger.

The final point is that participants employ a trend-following forecasting rule only when $\beta > 0$. Also, in this case, if there was inflation in the past, they predict higher prices in the future: the coefficient at $j = 0$ is 0.5, but at $j = 4$, it is 1.2, which is significantly higher. In contrast, with $\beta < 0$, participants scarcely rely on past inflation. This stark difference is thought to arise from whether the economy is

Table 3: Regression Results

| **Panel A** ($\beta = 0.9$) | Horizon $j = 0$ | $j = 1$ | $j = 2$ | $j = 3$ | $j = 4$ |
|---|---|---|---|---|---|
| $\Delta\alpha_t$ | 0.789 | 1.084** | 1.212** | 2.917** | 3.325** |
| | (0.537) | (0.479) | (0.617) | (1.186) | (1.386) |
| $1_{A_t}$ | 1.405 | 2.100* | 2.949** | 5.060* | 4.457 |
| | (1.212) | (1.081) | (1.393) | (2.679) | (3.130) |
| $\gamma_j \omega_j$ | 0.301*** | 0.255*** | 0.197*** | 0.244*** | 0.110*** |
| | (0.016) | (0.014) | (0.019) | (0.036) | (0.042) |
| $(1 - \gamma_j)\chi_j$ | 0.518*** | 0.644*** | 0.888*** | 1.331*** | 1.249*** |
| | (0.054) | (0.048) | (0.062) | (0.120) | (0.140) |
| $N$ | 3,168 | 3,168 | 3,168 | 3,168 | 3,168 |
| $R^2$ | 0.162 | 0.186 | 0.117 | 0.129 | 0.174 |
| **Panel B** ($\beta = -0.9$) | Horizon $j = 0$ | $j = 1$ | $j = 2$ | $j = 3$ | $j = 4$ |
| $\Delta\alpha_t$ | 0.250*** | 0.257*** | 0.294*** | 0.303*** | 0.248*** |
| | (0.030) | (0.023) | (0.029) | (0.028) | (0.028) |
| $1_{A_t}$ | 3.553*** | 1.856* | 4.507*** | 4.830*** | 5.364*** |
| | (1.273) | (1.002) | (1.231) | (1.186) | (1.198) |
| $\gamma_j \omega_j$ | 0.370*** | 0.228*** | 0.265*** | 0.161*** | 0.185*** |
| | (0.015) | (0.011) | (0.014) | (0.014) | (0.014) |
| $(1 - \gamma_j)\chi_j$ | 0.213*** | 0.013 | 0.050 | −0.055 | −0.018 |
| | (0.047) | (0.037) | (0.045) | (0.043) | (0.044) |
| $N$ | 3,120 | 3,120 | 3,120 | 3,120 | 3,120 |
| $R^2$ | 0.244 | 0.212 | 0.201 | 0.128 | 0.116 |
| **Panel C** ($\beta = -1.8$) | Horizon $j = 0$ | $j = 1$ | $j = 2$ | $j = 3$ | $j = 4$ |
| $\Delta\alpha_t$ | 0.183*** | 0.197*** | 0.173*** | 0.207*** | 0.204*** |
| | (0.032) | (0.031) | (0.038) | (0.037) | (0.041) |
| $1_{A_t}$ | 3.280 | 3.290* | 7.498*** | 8.841*** | 6.550** |
| | (1.999) | (1.945) | (2.427) | (2.345) | (2.572) |
| $\gamma_j \omega_j$ | 0.331*** | 0.292*** | 0.303*** | 0.286*** | 0.216*** |
| | (0.017) | (0.016) | (0.021) | (0.020) | (0.022) |
| $(1 - \gamma_j)\chi_j$ | 0.015 | 0.005 | −0.005 | −0.022 | −0.108** |
| | (0.041) | (0.040) | (0.050) | (0.048) | (0.053) |
| $N$ | 2,964 | 2,964 | 2,964 | 2,964 | 2,964 |
| $R^2$ | 0.270 | 0.257 | 0.189 | 0.617 | 0.575 |

converging to the steady-state REE prices or not. When $\beta > 0$, the economy deviates from the REE prices persistently, and the realized prices move in a nonstationary manner. Therefore, forecasting prices becomes more difficult when $\beta > 0$. This would be a reason why participants also use past information to predict prices. If $\beta < 0$, the past information about prices becomes less relevant (because prices fluctuate around the steady-state REE prices.)

Our results in this section can be summarized as follows. There exists a group of participants who consistently stick to naive expectations, even when faced with dynamic and evolving scenarios. In contrast, a significant number rely on both the eductive and evolutive methods of expectation formation, as implied by Binmore (1987). Importantly, there does not exist *the* ideal model for nonrational expectations formation. The appropriate approach for modeling expectations largely hinges on prevailing economic conditions that participants face. Specific factors, such as the degree of strategic complementarity shown by $\beta$, the presence of shocks, $\Delta\alpha_t$, and their announcement, play a pivotal role in influencing these expectation formations. Thus, appropriate models should allow endogenous selection of an optimal economic forecasting model tailored to the environment that participants face. Papers consistent with our findings include Anufriev and Hommes (2012) and Evans et al. (2021). Extending this line of research would be promising.

# 6 Concluding Remarks

In this paper, we extend the conventional single-period beauty contest game to a multi-period context and conduct experiments. This multi-period model is constructed based on the Calvo (1983) pricing model commonly used in macroeconomics. Our new paradigm of "learning-to-forecast" extends the existing ones and allows us to investigate how participants update their forecasts for the same future periods over time in response to the arrival of new pieces of information. The experimental results indicate that, on average, individuals react to announcements of future shocks (eductive reasoning), while a sizable fraction of participants forms naive expectations even in the multi-period environment considered in our experiment. Additionally, the incorporation of trends, shocks, and their announcement into expectation formation varies depending on economic conditions. Fitting the model in which decision makers choose different forecasting models depending on the situation, as in Anufriev and Hommes (2012) and Evans et al. (2021), to our data would be a fruitful next step.

Another approach is to conduct experiments to understand how participants select the variables to use for forecasting, and under which circumstances they are more susceptible to shocks. This is important because changes in the way people form expectations can impact the dynamics of the economy and potentially cause significant fluctuations through expectational feedback loops.

To gain a deeper understanding of the above point, it might be beneficial to change the magnitude of the shocks. When a game is with strategic complementary, even a small shock is magnified, exerting a considerable influence on equilibrium prices. However, if individuals expect that others do not respond to the shock, its initial impact on the economy is tiny, and the economy will slowly transition to the new equilibrium price. This is markedly different from the case of strategic substitutes. In that case, if people expect that others do not react to the shock, it can have a bigger impact on the economy, which might explain why participants respond to the shocks in games with strategic substitutes. To explore the impact of shocks and their announcement, it is helpful to consider sufficiently large shocks in experiments of games with strategic complementarity. These remain as future research topics.

# References

**Adam, Klaus**, "Experimental evidence on the persistence of output and inflation," *Economic Journal*, 2007, *117* (603-636).

**Afrouzi, Hassan, Spencer Y. Kwon, Augustin Landier, Yueran Ma, and David Thesmar**, "Overreaction in expectations: Evidence and theory," *Quarterly Journal of Economics*, 2023, *138* (3), 1713–1764.

**Akiyama, Eizo, Nobuyuki Hanaki, and Ryuichiro Ishikawa**, "How do experienced traders respond to inflows of inexperienced traders? An experimental analysis," *Journal of Economic Dynamics and Control*, 2014, *45*, 1–18.

**_ , _ , and _** , "It is not just confusion! Strategic uncertainty in an experimental asset market," *Economic Journal*, 2017, *127*, F563–F580.

**Angeletos, George Marios and Chen Lian**, "Forward guidance without common knowledge," *American Economic Review*, 2018, *108*, 2477–2512.

**Anufriev, Mikhail, Aleksei Chernulich, and Jan Tuinstra**, "Asset price volatility and investment horizons: An experimental investigation," *Journal of Economic Behavior and Organization*, 2022, *193*, 19–48.

**_ and Cars Hommes**, "Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments," *American Economics Journal, Microeconomics*, 2012, *4* (4), 35–64.

**_ , John Duffy, and Valentyn Panchenko**, "Learning in two-dimensional beauty contest games: Theory and experimental evidence," *Journal of Economic Theory*, 2022, *201*, 105417.

**Asparouhova, Elena, Peter Bossaerts, Nilanjan Roy, and William Zame**, "“Lucas” in the laboratory," *Journal of Finance*, 2016, *71* (6), 2727–2779.

**Assenza, T., P. Heemeijer, C.H. Hommes, and D. Massaro**, "Managing self-organization of expectations through monetary policy: A macro experiment," *Journal of Monetary Economics*, 2021, *117*, 170–186.

**Bao, Te, Cars Hommes, and Jiaoying Pei**, "Expectation formation in finance and macroeconomics: A review of new experimental evidence," *Journal of Behavioral and Experimental Finance*, 2021, *32*, 100591.

**_ , _ , and Tomasz Makarewicz**, "Bubble foremation and (in)efficient markets in learning-to-forecast and -optimize experiments," *Economic Journal*, 2017, *127*, F581–F609.

**_ , _ , Joep Sonnemans, and Jan Tuinstra**, "Individual expectations, limited rationality and aggregate outcomes," *Journal of Economic Dynamics and Control*, 2012, *36*, 1101–1120.

**_ , John Duffy, and Cars Hommes**, "Learning, Forecasting, and Optimizing: An Experimental Study," *European Economic Review*, 2013, *61*, 186–204.

**_ , Myrna Hennequin, Cars Hommes, and Domenico Massaro**, "Coordination on bubbles in large-group asset pricing experiments," *Journal of Economic Dynamics and Control*, 2020, *110*, 103702.

**Binmore, Ken**, "Modeling Rational Players: Part I," *Economics and Philosophy*, 1987, *3*, 179–214.

**Bosch-Domènech, Antoni and Nicolaas J. Vriend**, "Imitation of Successful Behaviour in Cournot Markets," *Economic Journal*, 2003, *113*, 495–524.

**Calvo, Guillermo A**, "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, 1983, *12*, 383–398.

**Chen, Daniel L., Martin Schonger, and Chris Wicken**, "oTree - An open-source platform for laboratory, online and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

**Coibion, Olivier, Yuriy Gorodnichenko, and Kamdar Saten**, "How do firms form their expectations? New survey evidence," *American Economic Review*, 2018, *108* (9), 2671–2713.

**Colassante, Annarita, Simone Alfarano, Eva Camacho-Cuena, and Mauro Gallegati**, "Long-run expectations in a learning-to-forecast experiment: a simulation approach," *Journal of Evolutionary Economics*, 2020, *30* (75-116).

**Crockett, Sean, John Duffy, and Yehuda Izhakian**, "An experimental test of the Lucas asset pricing model," *Reivew of Economic Studies*, 2019, *86*, 627–667.

**Evans, George W., Cars Hommes, Bruce McGough, and Isabelle Salle**, "Are Long-horizon Expectations (De-) stabilizing? Theory and Experiments," *Journal of Monetary Economics*, 2022, *132*, 44–63.

_ **, Christopher G. Gibbs, and Bruce McGough**, "A unified model of learning to forecast," mimeo, University of Oregon 2021.

**Farhi, Emmanuel and Ivan Werning**, "Monetary Policy, Bounded Rationality, and Incomplete Market," *American Economic Review*, 2019, *109*, 3887–3928.

**Fehr, Ernst and Jean-Robert Tyran**, "Limited Rationality and Strategic Interaction: The Impact of the strategic environment on nominal inertia," *Econometrica*, 2008, *76* (2), 353–394.

**Fréchette, Guillaume R. and Sevgi Yuksel**, "Infinitely repeated games in the laboratory: four perspectives on discounting and random termination," *Experimental Economics*, 2017, *20*, 279–308.

**Funaki, Yukihiko, Kohei Kawamura, Nobuyuki Uto, and Kozo Ueda**, "Price Stickiness and Strategic Uncertainty: An Experimental Study," CAMA Working Paper 63/2023, Crawford School of Public Policy, Australian National University 2023.

**Galí, Jordi**, *Monetary policy, inflation and the business cycle*, 2nd edition ed., Princeton University Press, 2015.

**García-Schmidt, Mariana and Michael Woodford**, "Are Low Interest Rates Deflationary? A Paradox of Perfect-foresight Analysis," *American Economic Review*, 1 2019, *109*, 86–120.

**Greiner, Ben**, "An Online Recruitment System for Economic Experiments," *Journal of the Economic Science Association*, 2015, *1* (1), 114–125.

**Hanaki, Nobuyuki, Yukio Koriyama, Angela Sutan, and Marc Willanger**, "The strategic environment effect in beauty contest games," *Games and Economic Behavior*, 2019, *113*, 587–610.

**Haruvy, Ernan, Yaron Lahav, and Charles N. Noussair**, "Traders' Expectations in Asset Markets: Experimental Evidence," *American Economics Review*, December 2007, *97* (5), 1901–1920.

**Heemeijer, Peter, Cars Hommes, Joep Sonnemans, and Jan Tuinstra**, "Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation," *Journal of Economic Dynamics and Control*, 2009, *33*, 1052–1072.

**Hommes, Cars, Anita Kopányi-Peuker, and Joep Sonnemans**, "Bobbles, crashes and information contagion in large-group asset market experiments," *Experimental Economics*, 2021, *24*, 414–433.

__ **, Joep Sonnemans, Jan Tuinstra, and Henk van de Velden**, "Coordination of expectations in asset pricing experiments," *Review of Financial Studies*, 2005, *18* (3), 955–980.

**Huck, Steffen, Hans-Theo Normann, and Jörg Oechssler**, "Learning in Cournot Oligopoly - An Experiment," *Economic Journal*, 1999, *109* (454), C80–C95.

__ **,** __ **, and** __ , "Two are few and four are many: number effects in experimental oligopolies," *Journal of Economic Behavior and Organization*, 2004, *53*, 435–446.

**Kryvtsov, Oleksiy and Lube Petersen**, "Central bank communication that works: Lessons from lab experiments," *Journal of Monetary Economics*, 2021, *117*, 760–780.

**Lucas, Robert E.**, "Asset prices in an exchange economy," *Econometrica*, 1978, *46*, 1429–1445.

**Lustenhouwer, Joep and Isabelle Salle**, "Forecast revisions in the presense of news: a lab investigation," AWI Discussoin Paper 714, Department of Economics, Heidelberg University 2022.

**Marimon, Ramon and Shyam Sunder**, "Indeterminacy of equilibria in a hyperinflationary world: Experimental evidence," *Econometrica*, 1993, *61* (5), 1073–1107.

**Mokhtarzadeh, Fatemeh and Luber Petersen**, "Coordinating expectations through central bank projections," *Experimental Economics*, 2021, *24*, 883–918.

**Noussair, Charles N., Damjan Pfajfar, and Janos Zsiros**, "Pricing decisions in an experimental dynamic stochastic general equilibrium economy," *Journal of Economic Behavior and Organization*, 2015, *109*, 188–205.

**Offerman, Theo, Jan Potters, and Joep Sonnemans**, "Imitation and belief learning in an oligopoly experiment," *Review of Economic Studies*, 2002, *69*, 973–997.

**Orland, Andreas and Michael W.M. Roos**, "The New Keynesian Phillips curve with myopic agents," *Journal of Economic Dynamics and Control*, 2013, *37*, 2270–2286.

_ **and** _ , "Price-setting with quadratic adjustment costs: Experimental evidence," *Journal of Economic Behavior and Organization*, 2019, *163*, 88–116.

**Petersen, Luba**, "Do expectations and decisions respond to monetary policy?," *Journal of Economic Studies*, 2015, *42* (6), 972–1004.

_ **and Ryan Rhoes**, "Macroeconomic expectations, central bank communication, and background uncertainty: A COVID-19 laboratory experiment," *Journal of Economic Dynamics and Control*, 2022, *143*, 104460.

**Potters, Jan and Sigrid Suetens**, "Cooperation in experimental games of strategic complements and subsitutes," *Review of Economic Studies*, 2009, *76*, 1125–1147.

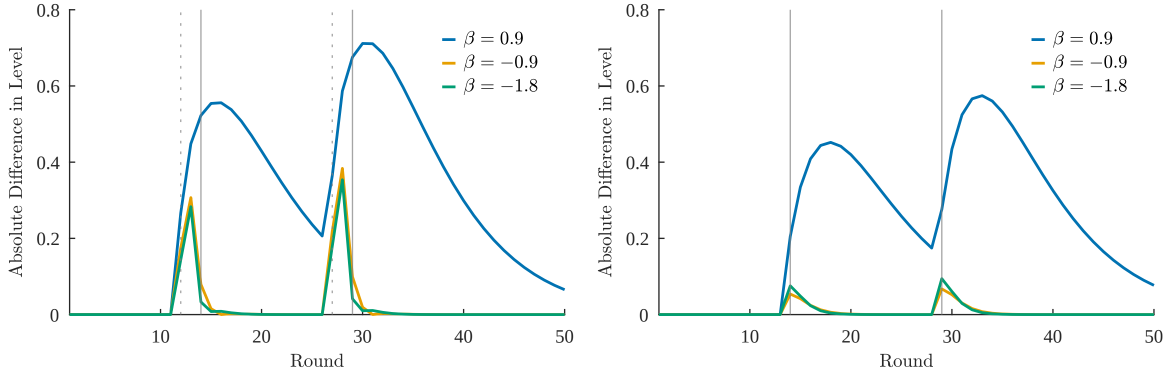
**Rholes, Ryan and Luba Petersen**, "Should central banks communicate uncertainty in their projections?," *Journal of Economic Behavior and Organization*, 2021, *183*, 320–341.

**Smith, Vernon L., Gerry L. Suchanek, and Arlington W. Williams**, "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, 1988, *56* (5), 1119–1151.

**Stöckl, Thomas, Jürgen Huber, and Michael Kirchler**, "Bubble measures in Experimental Asset Markets," *Experimental Economics*, 2010, *13* (3), 284–298.

**Sutan, Angela and Marc Willinger**, "Guessing with negative feedback: An experiment," *Journal of Economic Dynamics and Control*, 2009, *33* (5), 1123–1133.

**Woodford, Michael**, *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, 2003.




(a) With Announcement

(b) Without Announcement

Figure 6: Comparison

# A Comparison of REE Prices for $\gamma > 0$ and $\gamma = 0$

In our experiments, we inadvertently ignored the impacts of termination probability, $\gamma$, in determining the price. However, this oversight is not likely to result in significant changes to our study. This is because the impacts are minor as mentioned in footnote 11. In this subsection, we formally substantiate our claim. We calculate the REE prices under the assumption that $\gamma = 5\%$, and compare the prices with those computed under the assumption that $\gamma = 0$. We find that the prices are sufficiently similar between the two.

Let $(p_t^\gamma, P_t^\gamma)$ denote the equilibrium individual price and the equilibrium price when $\gamma \geq 0$.

Figure 6 depicts the absolute difference between $P_t^{0.05}$ and $P_t^0$ in case with and without announcement. Note that the differences are less than one in all the periods including those around the shocks.

# B Detail of the Procedure of the Online Experiment

Participants join our experiments via Zoom with their cameras and microphone turned off. The camera of the experimenter is always turned on, but her or his microphone is turned on only when necessary.

Upon connecting to a Zoom session, participants first wait in the waiting room. We let participants enter the main room one by one to check their names and to verify whether they are indeed registered for our experiments. Then, each participant is given participant ID in the form of "sub##", where ## is the two-digit number that is valid during the experiment. Once their participant ID is given, they are sent back to the waiting room until the start of the experiment. By following this procedure, we ensure anonymity.

Once ready, participants re-enter the Zoom meeting room and are given general instructions regarding the online experiment (for example, what to do, including which number to call when their internet connection fails during the experiment). Then, the prerecorded instruction video is played first. Although participants are not given a hard copy of the instruction slides, they are informed that they can go through the same set of slides after the video finishes until they finish answering the comprehension quiz. All the participants need to answer all six questions of the quiz correctly for the first game to start. As noted, participants can review the instruction slides before and while answering the quiz. While participants are asked to communicate directly with the experimenter using the chat function of Zoom when they have questions or encounter problems, they could not communicate with each other via Zoom chat.

# C English Translations of the Instructions and Examples of Screenshots

English translation of the instruction slides can be found at https://osf.io/gecv8/?view_only=005886a62c3148f6afe46f0406b2887d

Figures 7 to 10 show examples of the decision screen participants faced in our experiment.


Note: The values of $\alpha$ and $\beta$ are shown in red and with a large font in period 1 as well as when they change (in periods 14 and 29) in all the treatments (see Figure 9). In other periods, they are shown in black with a regular sized font (see Figure 10).

Figure 7: Screen (in period 1) in which participants submit their five forecasts (common to all the treatments)


Note: This is a demo screen in which the shock is introduced in period 5 (this is why the pre-announcement is made in period 3). The text in the yellow box states that "This is information regarding the future changes in the parameter values.
The values of the parameters that determine the price will be
$\alpha = xxx$ and $\beta = yyy$ from period $T$," where $xxx$, $yyy$, and $T$ depend on the treatment and whether it is the first shock or the second shock. The current values of $\alpha$ and $\beta$ are shown in black text below the announcement.

Figure 8: Screen with a pre-announcement of future shock (only for the treatments with pre-announcement)


Note: The values of $\alpha$ and $\beta$ are shown in red with a large font when they change (in periods 14 and 29) in all the treatments.

Figure 9: Screen when the shock is realized (common to all the treatments)


Note: The values of $\alpha$ and $\beta$ are shown in black with a regular sized font.

Figure 10: Screen for normal periods (common to all the treatments)

## C.1 Comprehension quiz

Here is an English translation of comprehension quiz.

(1) In this experiment, every morning, you will be asked to predict prices up to K periods ahead, including the current period. For the forecasts that will be rewarded in period $t$ ($t > 1$), please select all correct statements.

   A) The latest price forecast entered in period $t$ will be the subject of the reward.

   B) There is a 0.5 probability that the latest price forecast entered in period $t$ will be the subject of the reward.

   C) If the latest price forecast entered in period $t$ is not the subject of the reward, then the price forecasts entered in period $t-1$ or earlier will be the subject of the reward.

(2) Let's assume the price forecast that is the subject of the reward in period $t$ is 10. Also, let's assume the price that materialized in period $t$ is 14. In this case, how many points can you earn?

   A) 20 points.

   B) 25 points.

   C) 50 points.

(3) Regarding how the price in period $t$ is determined, which statement is correct?

   A) The realized price is determined based on the latest price forecasts submitted by participants in the same group for period $t$.

B) For participants in the same group, the realized price is determined based on the payoff-relevant price forecast in period $t$.

(4) Please select all correct statements about the repetition period of a single game.

A) A game is repeated for at least 20 periods.

B) The game can end after just one period as there is a 0.05 probability of the game ending at the end of each period.

C) The game is repeated only for 20 periods.

(5) Which statement is correct regarding the points that can be earned in a single game?

A) Since the game is repeated for at least 20 periods, you can earn all the points accumulated during at least 20 periods.

B) Although the game is repeated for at least 20 periods, the number of points that can be earned might be less than the total points accumulated over 20 periods, because there is a probability of 0.05 that the game will end at the end of each period.

(6) Please select all correct statements regarding today's experiment.

A) The game is conducted only once, and the compensation is paid according to the points earned in that game.

B) It is not known in advance how many times the game will be conducted.

C) One of the games conducted will be randomly selected to be the subject for compensation.