Discussion Paper No. 1213

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

AN EXPERIMENT ON A MULTI-PERIOD BEAUTY CONTEST GAME

Nobuyuki Hanaki
Yuta Takahashi

Secondly Revised December 2024
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‡

December 23, 2024

Abstract

We present and conduct a novel experiment on a multi-period beauty contest game. Leveraging the multi-period feature, we propose a new methodology to test the forward-lookingness of expectations by studying how expectations are revised over time. Our experimental results show that expectation formation is indeed forward-looking. Moreover, we uncover a new effect of strategic environment by exploring how expectations are formed in our dynamic environment: only when the game exhibits strategic complementarity do participants use extrapolation and expect increasingly higher prices in the future. This finding implies that the mode of expectation formation is endogenous to the economic environment of the participants.

Keywords: Expectation formation, Learning-to-forecast experiment

JEL Code: C92, D84, E70

*We thank Elena Asparouhova, Te Bao, John Duffy, Rosemarie Nagel, Charles Noussair, Yoshiyasu Ono, Lube Petersen, Marc Willinger, and the participants at the 2023 CEFM workshop (Nagasaki), AP-ESA 2023 (Seoul), EF 2023 (Sofia), 2023 ESA World Meeting (Lyon), and 2024 Experimental Macro Workshop (Singapore) 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 the 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

Rational expectations have long been a cornerstone in macroeconomic modeling. Over the last five decades, a substantial body of macroeconomic 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).

Experimental evidence from "learning-to-forecast" experiments further questions the validity of the REE hypothesis. These experiments, pioneered by Marimon and Sunder (1993), reveal that participants' behaviors often diverge from the rational expectations benchmark, especially when the aggregate outcome is determined endogenously by individual forecasts (Hommes et al., 2005; Heemeijer et al., 2009; Bao et al., 2017).[1] The presence of *expectational feedback* – where individual expectations influence the aggregate outcome, which in turn shapes future expectations – can lead to dynamics that are better captured by models incorporating backward-looking expectation formation (Anufriev and Hommes, 2012; Evans et al., Forthcoming).

However, most existing learning-to-forecast experiments are essentially of the single-period type.[2] Participants only need to forecast the outcome of today or in one future

[1] Note that unlike experiments in which participants actually trade in the market (e.g., Smith et al., 1988; Haruvy et al., 2007; Akiyama et al., 2017; Asparouhova et al., 2016; Crockett et al., 2019), set prices (e.g., Fehr and Tyran, 2008; Noussair et al., 2015; Orland and Roos, 2013, 2019; Petersen, 2015), or set quantities (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; they 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 between learning-to-forecast experiments and 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 (e.g., Kryvtsov and Petersen, 2021; Mokhtarzadeh and Petersen, 2021).

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

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 and policymakers who need to know how agents form expectations to design better policies in a dynamic world. In typical macroeconomic models and dynamic games more generally, 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 current outcomes. The single-period framework does not enable us to analyze how these series of expectations change in such a multi-period environment, such as in response to news about future shocks.

In order to fill this gap in the literature, we propose and conduct a novel experiment of a multi-period beauty contest game where participants submit a sequence of forecasts for multiple future periods. Building on this new experimental design, we make two main contributions. First, we propose a new methodology to test the forward-lookingness of expectations. Second, we explore how individuals form their expectations in a multi-period beauty contest game.

Our experiment is motivated by the core component of the canonical New-Keynesian model (e.g., Woodford (2003) and Galí (2015), 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 a given firm depends on its own price and the aggregate price level. While an increase in the firm's own price reduces the demand for that firm, demand may increase or decrease when other firms raise their prices. Suppose for a moment that firms can set their prices freely. In this case, they choose prices to maximize profit, taking other firms' prices as given. An REE price is a fixed point in this process and is often referred to as the optimal flexible price.

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 (e.g., Coibion et al., 2018).

Following Calvo (1983), we now introduce a friction that limits firms' ability to choose an optimal flexible price. Instead, firms can only change their prices with a probability of $1 - \theta$. This constraint induces them to take future prices into account because future profit depends on current price choices; if prices cannot be adjusted later, firms must compete in future periods with the price selected today. Hence, they must incorporate future market conditions and price expectations into their current decisions. This mechanism introduces an additional dynamic feedback effect, since beliefs about future prices influence today's actions.

We transform our theoretical model into a multi-period version of the "learning-to-forecast" game as in Marimon and Sunder (1993). In each period, participants submit forecasts for prices in multiple future periods all at once. These forecasts determine not only participants' payoffs but also the realized aggregate prices. Following the literature on learning-to-forecast game experiments, participants merely need to anticipate aggregate prices accurately. Their payoffs are maximized when they submit accurate forecasts each period, regardless of whether the game features strategic substitutes or complements.

We adopt a three-by-two between-subjects experimental design that varies two key dimensions. First, we consider the strategic environment, examining treatments that feature strong strategic complementarity, and those with equally strong and even stronger degrees of substitutions. Second, we manipulate the timing of information about upward shifts in the demand curve which we call shocks. In some treatments, participants face these shocks unexpectedly; in others, they are announced two periods in advance. This design allows us to investigate both how the strategic environment affects behavior and whether participants form forward-looking expectations in response to anticipated shocks.

We begin by examining the effects of the strategic environment. This analysis serves as a check to ensure that our findings align with those of existing one-period learning-to-forecast games, rather than as a novel contribution.[4] We find that well-documented patterns

[4]Existing learning-to-forecast experiments, such as Heemeijer et al. (2009) and Bao et al. (2017), demonstrate that, on one hand, prices deviate substantially from those expected under REE when there is a strong

of existing papers are replicated within our new experimental framework, thereby validating the internal consistency of our setup before we move on to more substantive analyses.

Having replicated the known results of the strategic environment within our experimental framework, we next examine the forward-looking nature of expectation formation. A key innovation of our design is that participants submit a sequence of forecasts extending into future periods, including those after an announced shock has occurred. For example, when a shock scheduled two periods ahead is announced in advance, we can analyze how participants revise their forecasts following the shock's realization. This feature highlights the novelty of our approach; rather than relying solely on forecasts made for the immediate next period, we explicitly capture adjustments that unfold after anticipated events have materialized.

Individual responses to these announcements exhibit considerable heterogeneity. While some participants anticipate prices that exceed the levels predicted by the REE, others appear to disregard the announcements altogether. Such diversity in behavior underscores the importance of focusing on average responses to determine whether, despite heterogeneity, participants as a whole adjust their forecasts in a manner consistent with forward-looking reasoning.

Our analysis demonstrates that, on average, participants do respond to the announcements in line with forward-looking behavior. This finding provides new evidence of forward-lookingness in expectation formation, offering a distinctive contribution to the literature.

Having established that participants exhibit forward-lookingness in their forecast revisions, we shift our focus from average patterns to the finer details of individual expectation formation. Specifically, we turn to periods absent of shocks and their announcements to

positive feedback between expectations and prices and on the other hand, 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; Cooper et al., 2021; 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) have demonstrated 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 multi-period framework.

investigate how participants form beliefs in a dynamic environment under "normal" conditions.[5] For this objective, we estimate a reduced-form forecasting rule that encompasses various heuristics like adaptive expectations and trend following, allowing coefficients to vary across different forecast horizons.

The regression results show that most participants rely heavily on the most recent price as a reference point, with coefficients close to one, and they also consider their past forecast errors, indicating self-referential expectations. Notably, the influence of recent price *changes* becomes more pronounced as the forecast horizon extends, especially in positive feedback settings, where participants increasingly depend on trend-following behavior for longer-term forecasts. This finding suggests that in positive feedback environments, participants expect price increases to persist, potentially leading to de-anchored long-run inflation expectations. Conversely, in negative feedback settings, participants view price increases as temporary, and their expectations remain relatively stable across horizons.

Taken together, these findings indicate an important policy implication: expectation formation depends qualitatively on whether the strategic environment features complementarity or substitutability. As a result, central banks must understand the strategic context in which agents are operating to effectively conduct monetary policy and mitigate the risks of expectation de-anchoring.

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.[6] 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's asset pricing model (1978). Participants

[5] Unfortunately, it is difficult to study how each individual participant responded to the shocks and their announcements since we only have two observations for each participant.

[6] 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 not only forecast future prices but actually trade the asset; see, e.g., Haruvy et al. (2007) and Akiyama et al. (2014, 2017).

submit their forecast for *the average price over multiple future periods*, instead of the forecasts about the prices for each of these future periods, which determines the prices. They mix 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 with only short-horizon forecasters exhibit substantial and prolonged deviation from the REE, those markets with even a modest share of long-horizon forecasters converge.

Adam (2007), Rholes and Petersen (2021), and Petersen and Rhoes (2022) 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 current outcomes. 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) and 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, and 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, those authors examine the impact of the announcement regarding future policy changes on the expectations, as we do in our experiment. In their study, participants receive only qualitative information regarding how their forecasts, along with policy variables, collectively influence the output. In our study, however, participants are informed about how their forecasts collectively affect the outcomes.

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 experiment's design and procedure are presented in Section 3, while 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 theoretical model, which serves as the foundation for our experiment design. Consider a continuum of monopolistically competitive firms uniformly distributed over the interval $[0, 1]$. Each firm i selects a price p_i; for simplicity, we omit the subscript i. The demand function for an individual firm is

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

where P represents the aggregate price given by

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

We assume that $a > 0$, $b > 0$, and $c \in \mathbb{R}$, with the condition

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

which ensures the existence of an REE price in the single-period version of the game. The parameter c can be positive or negative, which governs the degree of strategic interaction among firms. All firms share an identical linear technology function with a unit production cost denoted by κ.

Each firm maximizes the present value of its profit but faces pricing frictions; firms cannot adjust their prices every period and can only do so with probability $1 - \theta$. If a firm cannot change its price for T consecutive periods, it is allowed to adjust its price with certainty in the next period.[7] Additionally, with probability $\gamma \in [0, 1)$, all firms may be forced to exit the market. This assumption enables us to conduct experiments within a finite period that

[7]In New Keynesian models, T is set to infinity. However, in our experiment, we aim to ensure that participants can reset their prices. Therefore, we choose a finite T.

effectively replicate those of an infinite period (Duffy, 2017).

Firms that can reset their prices in period t solve the following optimization problem:

$$\max_{p_t} \quad \sum_{s=0}^{T-1} ((1-\gamma)\theta)^s \left(p_t - \kappa\right) D\left(p_t; P_{t+s}\right). \tag{3}$$

Recall that in every period, firm i can re-optimize its price with probability $1-\theta$. Since the continuation payoffs in these future events are not a function of p_t, they do not appear in objective function (3).

Now, we formally define a *rational expectation equilibrium*. A rational expectation equilibrium is a pair of prices (p_t, P_t) such that given the aggregate price $(P_t)_t$, p_t solves the maximization problem (3), and the consistency, $P_t = p_t$, holds for all t.

We establish the following proposition:

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

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

where

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

Moreover, the aggregate REE price evolves according to

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

Proof. Taking the first-order conditions of the maximization problem in equation (3), we obtain Eq. (4). To derive Eq. (5), recall that the aggregate price in Eq. (2) is given by the average of individual prices 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 Eq. (4). So, average price P_t satisfies Eq. (5). $\square$

Notably, $\alpha + \beta P_t$ represents the optimal price in period t, if the firms can reset their prices freely. This price is often referred to as the optimal flexible price. Therefore, the optimal price in our environment, as shown in Eq. (4), constitutes a weighted average of the optimal flexible prices. Eq. (4) captures the dynamic thought process within firms; the optimal price, p_t, depends on current actions by others, as well as future actions, represented by $(P_{t+s})_{s=0}^{T-1}$. Because of this inter-period interdependence, today's aggregate price is also influenced by expectations of future prices. This feature leads to the model being referred to as a *dynamic beauty contest model* (e.g., Angeletos and Lian, 2018).

By contrast, when firms can adjust prices without friction, $\theta = 0$, the dynamic aspect disappears, and firms focus solely on predicting other firms' current actions. Our proposed experiment, inspired by this model, differs from existing single-period experiments by requiring participants to anticipate not only others' actions in period t but also in subsequent periods $t + 1, \ldots, t + T - 1$.

3 Experimental Design

3.1 Setup

In our experimental setup, we employ groups of six participants.[8] These participants, indexed by i, engage in the game over multiple periods. At the beginning of each period, they submit their price forecasts for the next five periods including the current one; that is, $T = 5$.[9] For example, in the first period, participants submit their price forecasts for periods 1 to 5. In the second period, they provide forecasts for periods 2 to 6, and so on. We denote the forecast of the price in period k submitted in period t as $f^i_{t,k}$. While not all submitted forecasts impact the participants' rewards, as detailed below, all five forecasts have the potential to do so when submitted.

[8]One may consider a group 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 even with close to 100 participants (Hommes et al., 2021).

[9]All the forecasts are constrained to be integers.

Let π_t^i denote the reward of participant i in period t, as determined by

$$\pi_t^i = \frac{100}{|F_t^i - P_t| + 1},$$

where P_t is the realized price in period t, and F_t^i is the payoff-relevant forecast of participant i in period t, as defined below.[10] If the reward is not an integer, then it is rounded to the nearest integer.

In the first period t, the payoff-relevant forecast F_t^i is set to $F_1^i = f_{1,1}^i$. In period 2, the payoff-relevant forecast F_2^i is determined probabilistically: with probability $1-\theta$, the newly submitted set of forecasts become payoff-relevant, yielding $F_2^i = f_{2,2}^i$; with probability θ, the previous forecasts remain payoff-relevant, resulting in $F_2^i = f_{1,2}^i$. This process continues in subsequent periods. If the same set of forecasts has been payoff-relevant for five consecutive periods (i.e., a firm cannot reset its price for five consecutive periods), then the new set of forecasts submitted in the next period becomes payoff-relevant with certainty.

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. 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. As in Marimon and Sunder (1993) and Bao et al. (2017), 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. See Appendix A for a formal proof.

To illustrate how submitted forecasts determine the payoff-relevant forecasts, consider the hypothetical forecasts of participant i shown in Table 1. Each row lists the forecasts submitted in period t. For example, in period $t = 1$, the participant submits forecasts of the

[10] This way of rewarding forecast accuracy is also used in, for example, Adam (2007), Assenza et al. (2021), and Anufriev et al. (2022b).

Period t	Forecast Periods k $k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	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**

Table 1: Hypothetical Submitted Forecasts

prices for periods 1 to 5:

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

In period 1, this set of forecasts becomes payoff-relevant so that the payoff-relevant forecast is $F_1^i = 10$. In period 2, the participant submits new forecasts $f_{2,2}^i, \cdots, f_{2,6}^i$. Suppose that the new forecasts do not become payoff-relevant. Then, $F_2^i = f_{1,2}^i = 11$. If the new set of forecasts becomes payoff-relevant in periods 3 and 4, then the latest forecasts determine the payoff-relevant forecasts. Thus, $F_3^i = f_{3,3}^i = 9$ and $F_4^i = f_{4,4}^i = 13$.[11] The reward for the participant is determined by the difference between the payoff-relevant forecasts F_t^i and aggregate price P_t.

We now explain how the aggregate price is determined in our experiments. In our experiments, all participants submit their forecasts simultaneously every period, and these submitted forecasts jointly determine aggregate price, P_t. Based on the model in Section 2, the aggregate price is given by

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

where α and β are parameters of the model and specified later. Eq. (6) is the empirical-counterpart of Eq. (2). The first term on the right-hand side of Eq. (6) represents the prices of participants who are unable to adjust their prices. The second term corresponds to the prices of participants who can reset their prices. The mapping of submitted forecasts to the

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

optimal price is given by Eq. (4).

On the screen in which participants submit their forecasts, the values of α and β are presented clearly. On the same screen, participants are informed of the realized P_t and the payoff-relevant forecast F_t in all past periods. See Appendix D for the screenshots.

In our experiment, we set the reset probability θ to 50% and assume that the game ends with probability $\gamma = 0.05$ at the end of each period.[12] Participants are rewarded based on the total points they earn throughout the game. Moreover, we set $T = 5$. Eq. (6) can be used for providing justification for $T = 5$. Note that when $\theta = 1/2$ and $\gamma = 0.05$, the impact of $f^i_{t,t+4}$ on the optimal price is minimal. This is because the weight of $f^i_{t,t+4}$ 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.

While it is conceptually trivial to end a game stochastically, implementing such a probabilistic termination rule in a laboratory experiment poses challenges, because the game might end too soon to study the evolution of the forecasts in response to shocks, or the game might not end within the scheduled time for participants. To address these challenges, we use the block random termination method (Fréchette and Yuksel, 2017) commonly used in experiments involving indefinitely repeated games.

Under this method, participants play the game in blocks of B periods. During each block, the game proceeds without participants knowing whether or not it has ended. Only at the end of a block are participants informed if the game actually ended at some point during that block.[13] If the game has ended during the block, they receive the sum of their payoffs π^i_t up to the period when the game ended. For example, if the game actually ended in period τ where $\tau < B$, participants are rewarded based on $\sum_{t=1}^{\tau} \pi^i_t$. If the game has not

[12] Due to an oversight in the experiment, the price was determined with $\gamma = 0$ instead of $\gamma = 0.05$. The incorrect price determination equation was also communicated to participants in the instructions. As shown in Appendix B, however, this oversight results in the REE price differing only slightly (by less than 1). Furthermore, because P_t is rounded to the nearest integer, we believe this oversight does not affect the experimental outcomes.

[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.

ended, it continues into the next block of B periods. They are also informed that the game can continue beyond B periods; if that happens, they play the game for at least another B periods. In our experiment, we set $B = 20$, so participants are told they will play the game for at least 20 periods.

Furthermore, since game durations may differ among groups, participants are told that the multi-period game can be repeated if all groups complete a game within 30 minutes from the first game's start. When a new game starts, participants are randomly re-grouped into six (requiring everyone in the session to wait until all groups have finished), and they play with the same parameter values.[14] If the game is repeated multiple times, one of them will be randomly selected for payment.[15]

During the games, we introduce (a maximum of) two shocks to α, which affects the demand size, during the first two blocks of 20 periods (one shock in each block). The literature suggests that 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 introduced the first shock at the beginning of period 14 and the second shock, if reached, at period 29. We assume that the initial flexible REE price is 65, after which it becomes 85 and 110 after the first and the second shocks, respectively. We choose α to match these price levels for each value of β.

3.2 Treatments

Our three-by-two between-participants experiment focuses on altering two main aspects of the games. The first is the degree of strategic interaction. We consider the games where β takes values in $\{0.9, -0.9, -1.8\}$. That is, the games exhibit either strategic complemen-

[14] We choose random re-matching because repeating the game with the same parameters could lead to coordination based on prior outcomes.

[15] To minimize variation in experiment duration and participant payments across sessions, we followed the procedure of Duffy and Puzzello (2014, 2022) by predefining the random number sequence. Predefining it ensured that all sessions repeated two blocks and concluded after the first game. Consistent with Duffy and Puzzello (2014, 2022), this information was not disclosed to participants.

tarity (positive feedback) or substitution (negative feedback).[16] We also explicitly consider a case of strong substitution with $\beta = -1.8$. This is motivated by the fact that New Keynesian models may exhibit strong substitutability, as demonstrated by García-Schmidt and Woodford (2019).

The second is the announcement of the shocks to α. We examine treatments both with and without pre-announcement of shocks. In the treatment without pre-announcement, participants are informed of the new value of α only when the shocks occur – 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 α in period 12 for the first shock and in period 27 for the second shock. This announcement specification is desirable because announcing the shocks two periods in advance allows us to examine participants' forecasts for prices *after the shocks*, both before and after the announcements.[17]

3.3 Procedures

Our experiments were conducted online using oTree (Chen et al., 2016), an open source platform for web-based interactive tasks. Participants joined from their own locations instead of our physical laboratory. We used Zoom to manage and coordinate the experiments.[18]

Once participants had received general instructions about the online experiment and were prepared, the prerecorded instruction video was shown on their screen. Although they did not receive physical copies of the instruction slides, participants were informed that they could access the same slides after the video finished, up until they finished the comprehension quiz. To begin the first game, all participants had to correctly answer all six quiz questions. Final rewards were provided through Amazon Gift Cards (e-mail version).

[16] Because participants are restricted to submitting integer forecasts, multiple equilibria may arise in cases of strong strategic complementarity. Specifically, when $\beta = 0.9$, multiple equilibria exist. However, the set of equilibria is limited, and this multiplicity does not significantly impact our results. Therefore, we do not address this issue explicitly.

[17] With this specification, 75% $(= 1 - (0.5)^2)$ of participants can change their forecasts before the realization of the shocks.

[18] See Appendix C for details regarding how we conducted our online experiment.

We recruited participants, who were students at Osaka University, using ORSEE (Greiner, 2015). An English translation of the instruction slides, examples of the decision screens, and the comprehension quiz are provided in Appendix D.

4 Benchmark Analysis: Rational Expectations Equilibrium

We take rational expectations as our benchmark. As shown in Proposition 1, the REE aggregate price denoted by $\left(P_t^{REE}\right)$ is characterized by Eqs. (4) and (5) given the initial condition, P_{-1}. Unlike the theoretical analysis, parameter α is time-varying in our experiment. Thus, we need to generalize Proposition 1 to accommodate this case. It is straightforward to show that the REE price with time-varying α satisfies the following equation:

$$P_t^{REE} = (1-\theta) \sum_{k=t}^{T-t-1} \frac{((1-\gamma)\theta)^{k-t}}{\sum_{s=0}^{T}((1-\gamma)\theta)^s} \left(\alpha_{t,k} + \beta P_{t+k}^{REE}\right) + \theta P_{t-1}^{REE}, \tag{7}$$

where $\alpha_{t,k}$ is the expected value of α in period k from the period-t perspective. When the shocks are not announced, $\alpha_{t,k} = \alpha_{t,t}$ for all k and t. When the shocks occur, then $\alpha_{t,t}$ suddenly increases. Suppose that the shocks are announced. Then $\alpha_{t,k}$ changes when the shocks are announced, not when they occur. Again, the announcement of the shocks matters only if the game is a multi-period beauty contest game. If firms do not face any pricing frictions $\theta = 0$, then the announcement has no effect on the REE price.

Figure 1 shows the REE price sequence $\left(P_t^{REE}\right)_t$ when the shocks are announced two periods in advance (left panel) and when the shocks are not announced (right panel). These figures are intuitively understood. When the game demonstrates strategic complementarity ($\beta > 0$), the transition to new steady state equilibrium prices is slower due to Calvo pricing friction. Participants recognize that others may not adjust their prices swiftly because of the pricing friction. Since individual optimal prices are positively related to others' pricing decisions, participants prefer to adjust their prices slowly.

By contrast, when the game exhibits strategic substitution ($\beta < 0$), the mechanism

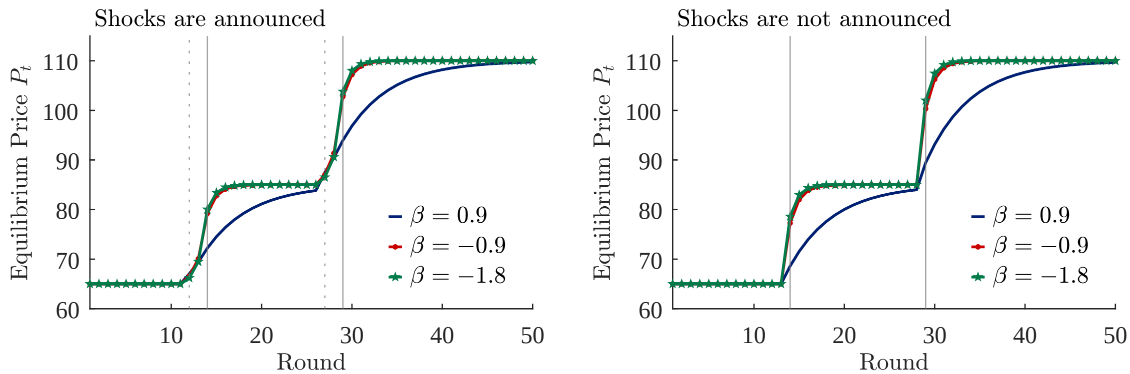


Figure 1: Aggregate Price P_t under Rational Expectation

Note: The solid vertical lines indicate when the two shocks occur, while the dotted lines in the left panel show when these shocks are announced.

operates in the opposite direction. If some agents fail to adjust their prices, the prevailing price becomes excessively low. This low price incentivizes agents to increase their prices. Consequently, the transition to the new steady state equilibrium price occurs more rapidly.

5 Results of the Experiment

We conducted our experiments in April and May 2023, involving a total of 294 participants.[19] Table 2 summarizes the number of groups per treatment. Each experiment lasted for approximately 90 minutes, and participants earned 2482 JPY (approximately 18 USD based on the exchange rate at the time), including a show-up fee of 500 JPY on average.[20] The average payment varied across the value of β. 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 $\beta = 0.9$,

[19]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 replaced this participant, allowing the experiment to continue. We exclude this group's data 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 or 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 prices 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.

[20]The exchange rate between points earned during the experiments and JPY was 1 point = 2 JPY.

Table 2: Number of Groups per Treatment

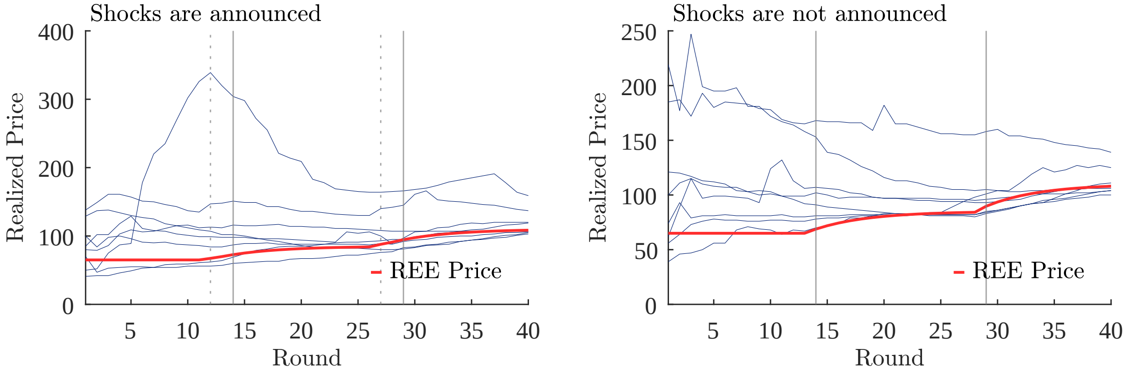
Treatments	$\beta = 0.9$	$\beta = -0.9$	$\beta = -1.8$
With Announcement	8	8^a	8
Without Announcement	8	8	7^b

Notes: "With Announcement" represents the results where the shocks were announced, while "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.

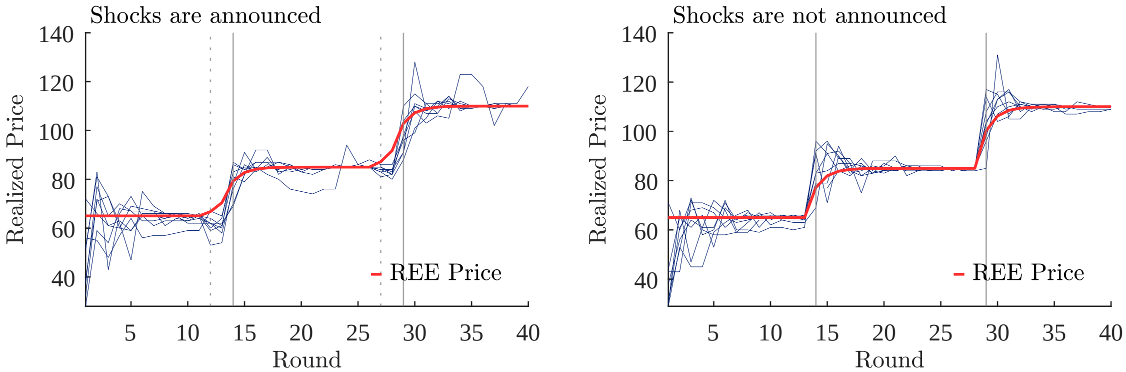
respectively.

5.1 Aggregate Price Dynamics

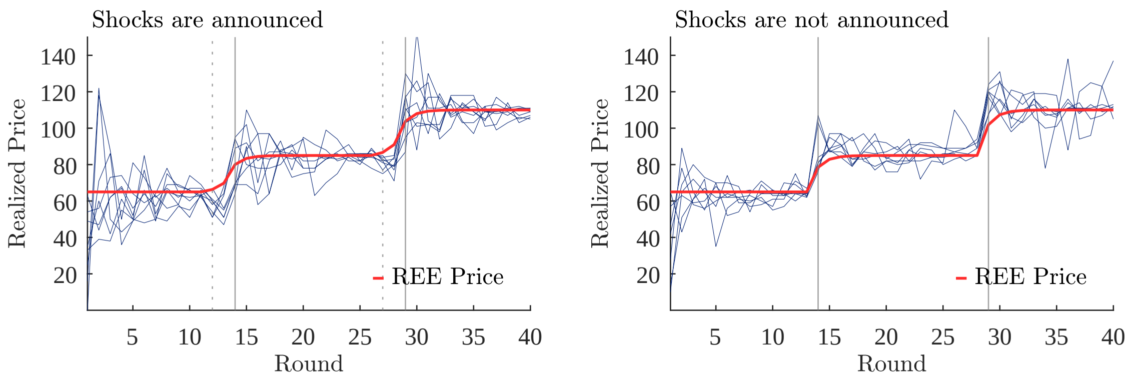
We begin by presenting the dynamics of prices observed in each treatment in Figure 2. Each line represents a group within each panel. As observed, irrespective of whether an announcement is made, the prices follow the REE price closely when the game exhibits strategic substitutes, $\beta < 0$. When the game exhibits strategic complements, $\beta > 0$, they persistently deviate from the REE prices. As shown by Fehr and Tyran (2008), these features can be understood intuitively. When the game exhibits strategic substitutes, the best response function has a positive slope, and participants have a strong motive to choose a similar price level of others. Consequently, the realized price remains close to the initial expectation of others' actions. Therefore, the initial expectation fulfills itself. This self-fulfilling mechanism makes the adjustment slow, leading to persistent deviations from the REE price. When the game exhibits strategic substitutes, then this mechanism does not operate. Participants want to choose higher (lower) prices when others choose lower (higher) prices. Thus, their initial expectation is not self-fulfilling unless it coincides with the REE price, and they often quickly converge to the REE price. Note that our experimental findings are also found in existing learning-to-forecast experiments, such as those by Heemeijer et al. (2009) and Bao et al. (2012).



(a) Positive Feedback: $\beta = 0.9$



(b) Weak Negative Feedback: $\beta = -0.9$



(c) Strong Negative Feedback: $\beta = -1.8$

Figure 2: Realized Aggregate Prices P_t

Notes: The red lines represents the aggregate price under the rational expectation. The solid vertical lines indicate when the two shocks occur, while the dotted lines in the left panel show when these shocks are announced.

It is noteworthy that aggregate prices become more stable when the game has weak strategic substitutes ($\beta = -0.9$). Conversely, with strong strategic substitutes, aggregate prices hover around the REE price but exhibit greater deviations. This phenomenon likely stems from difficulties in expectation coordination. When β is sufficiently negative, the realized price depends sensitively on expectations. This sensitivity results in the observed instability of prices.

5.2 Effect of the Degree of Strategic Interaction

We proceed to quantify the degree of deviation from the REE prices by following the methodology developed by Stöckl et al. (2010), which allows us to verify that our results align with evidence from existing one-period learning-to-forecast games. We compute the relative absolute deviation (RAD) and the relative deviation (RD) as proposed in their study. For each 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^{REE} is the REE price in period t.[21] K represents the total number of periods (40 except for the two groups that faced a technical problem).

Figure 3 shows the empirical cumulative distributions of RADs (top) and RDs (bottom) in the treatments with (left) and without (right) announcement. In each panel, the distributions for each β are shown. The top panels show that, regardless of the existence of announcement, RADs are positive for each β. Applying the signed-rank test to the distributions of RADs, the observed prices are significantly different from the REE prices.[22]

[21] 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 12.

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

Furthermore, the distributions of RADs are ranked in terms of first-order stochastic dominance; regardless of whether an announcement was made, the distribution under $\beta = 0.9$ stochastically dominates those under $\beta = -1.8$ and $\beta = -0.9$. This indicates a higher likelihood of larger deviations from the REE prices under $\beta = 0.9$ compared to the other treatments, and the differences between the three treatments are statistically significant both with and without announcement.[23] Therefore, deviations from the REE prices are smallest under weak substitution ($\beta = -0.9$), larger under strong substitution ($\beta = -1.8$), and largest under strategic complementarity ($\beta = 0.9$).

The bottom panels build on these findings by showing that RDs are also larger under strategic complementarity than under strategic substitution, mirroring the pattern observed for RADs. However, in contrast to the results for RADs, there is no longer a statistically significant difference between $\beta = -0.9$ and $\beta = -1.8$.[24] This statistical insignificance comes from the fact that RDs, by incorporating the direction of deviation from the REE price, become nearly zero when prices fluctuate around the REE price, as in the case where $\beta = -1.8$.[25]

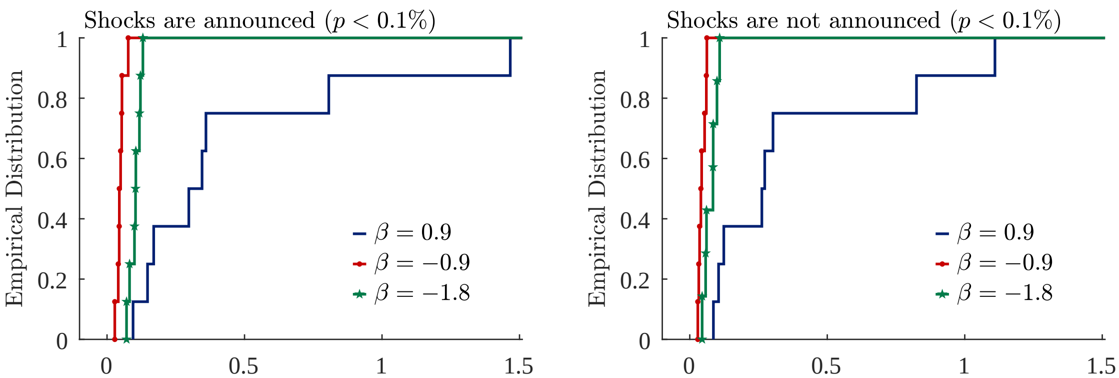
5.3 Forward-Lookingness of Expectations

We now examine whether forecasts have responded to the announcement of future shocks. To accomplish this objective, we analyze forecast revisions before and after those shock announcements. By studying these revisions, we can assess whether individuals have reacted to the announcements or disregarded them. In particular, if there are no revisions to forecasts

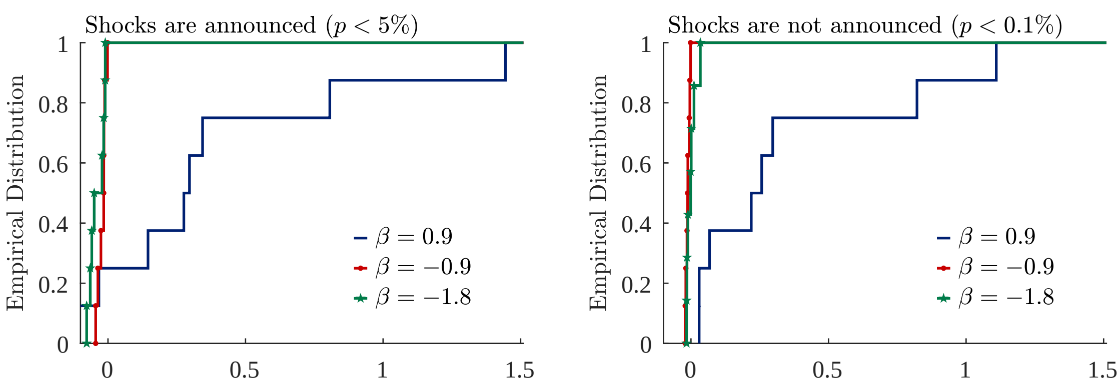
[23] $p < 0.05$ based on the Kruskal–Wallis test for both cases. For pairwise comparisons, p-values based on the two-tailed Mann-Whitney test are always less than 1%.

[24] For pairwise comparisons, p-values based on the two-tailed Mann-Whitney test 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.

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



(a) Relative Absolute Deviation (RAD)



(b) Relative Deviation (RD)

Figure 3: Deviations from the Rational Expectation Prices

Note: p-values of Kruskal-Wallis test are reported.

after an announcement, it suggests that the announcement was ignored.[26]

To operationalize this analysis, we define our measures of forecast revisions as follows. We focus on the forecasts for changes in prices during periods 14 and 29, $P_{14} - P_{13}$ and $P_{29} - P_{28}$, and analyze how these forecasts are revised before and after the associated shock announcements. These revisions are mathematically represented as

$$\Delta f_1^i = E^i \left[P_{14} - P_{13} \mid t \leq 13 \right] - E^i \left[P_{14} - P_{13} \mid t \leq 11 \right], \text{and} \tag{8}$$

$$\Delta f_2^i = E^i \left[P_{29} - P_{28} \mid t \leq 28 \right] - E^i \left[P_{29} - P_{28} \mid t \leq 26 \right], \tag{9}$$

where the conditional expectations are taken over the relevant information sets. For example, $E^i \left[P_{14} - P_{13} \mid t \leq 13 \right]$ represents the expected value of the change of the price in period 14 conditional on all information available up to period 13. Thus, the differences in the conditional expectations, Δf_1^i and Δf_2^i, capture the revisions of the forecasts of the price changes $P_{14} - P_{13}$ and $P_{29} - P_{28}$ in response to the announcement in period 12 and 27.[27] The conditional mean is calculated based on information available through periods 13 and 28 instead of periods 12 and 27, aiming to minimize noise by including two post-announcement observations.

By examining whether the values of these revisions are zero or not, we can infer that individuals are responding to the announcements. In particular, observing nonzero revisions strongly suggests that participants do react to the announcements. At the same time, because our analysis focuses on revisions tied to price changes, there is another way to interpret nonzero revisions: announcements exert a greater influence on forecasts for more distant future prices relative to the near term. [28] However, it is important to emphasize that identifying the forward-lookingness through these revisions is a sufficient but not strictly

[26]Note that this condition is sufficient but not necessary. As explained below, even if an individual revises her forecasts, she might do so for another reason and ignore the announcement.

[27]It is important to emphasize that we cannot compare Δf_1^i and Δf_2^i with the actual price changes in our data, $P_{14} - P_{13}$ and $P_{29} - P_{28}$. This is because Δf_1^i and Δf_2^i are *revisions* of the forecasts over $P_{14} - P_{13}$ and $P_{29} - P_{28}$, not the forecasts of these price changes.

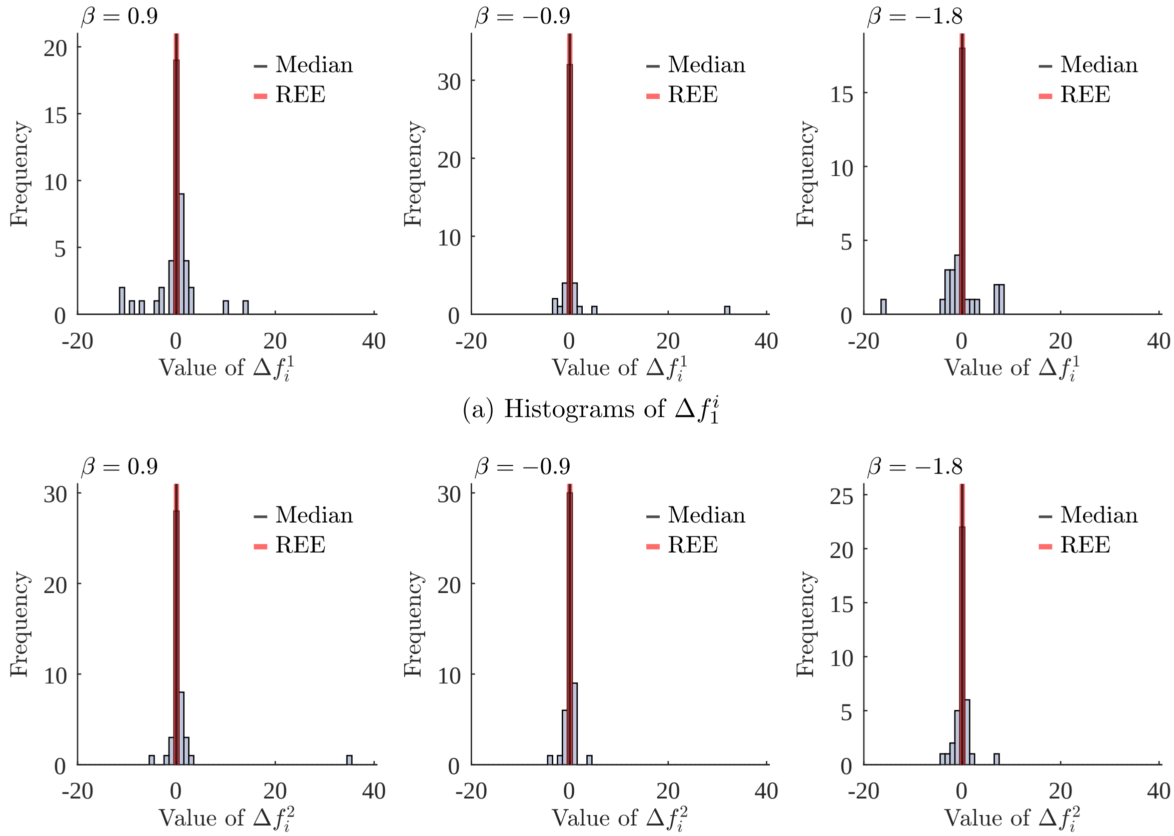
[28]Indeed, we can consider a revision of longer price change, defined as $\Delta f_1^i = E^i \left(P_{15} - P_{12} \mid t \leq 12 \right) - E^i \left(P_{15} - P_{12} \mid t \leq 11 \right)$. The analysis results below remain unchanged when using this measure.

necessary condition. For example, even if participant i revises her forecasts of both P_{14} and P_{13} by 10 points each in response to the first announcement, Δf_1^i remains at zero. Thus, according to our measure, this implies that she does not respond to the announcement. This example highlights that our inference based on the revisions provides a conservative gauge of the forward-lookingness. We intentionally adopt this conservative criterion to reduce the likelihood of falsely concluding that participants are forward-looking when they are not, thereby ensuring a more robust measure of genuine forward-looking behavior.

To calculate these revisions for each participant i, we take advantage of our experimental design. We measure the forecasts for the price change in period 14, conditional on the information available by period 13, as $\sum_{t=12}^{13}\left(f_{t,14}^i - f_{t,13}^i\right)/2$. Recall that $f_{t,14}^i - f_{t,13}^i$ represents the forecast of $P_{14} - P_{13}$ in period t. Hence, this average corresponds to the conditional expectation after the first shock announcement, if any. Similarly, we measure the conditional expectation before the first shock announcement, if any, as $\sum_{t=10}^{11}\left(f_{t,14}^i - f_{t,13}^i\right)/2$. The difference of these averages corresponds to Δf_1^i. We define Δf_2^i in an analogous way. Note that our multi-period experimental design uniquely allows us to measure these revisions. In the majority of existing papers, these revision measures are simply unavailable since sequences of forecasts are not collected.

To analyze the distributions of Δf_1^i and Δf_2^i, it is useful to compare them with their theoretical values under the REE. When the shocks are not announced, their theoretical values are easily obtained. Under the REE hypothesis, the entire path of the prices P_t is rationally expected at the beginning of the game. Thus, no additional information becomes available over time, and forecasts are not revised at all. Therefore, both Δf_1^i and Δf_2^i are zero in this case.

Consider the case where the shocks are announced. Since the announcements of the shocks are new information, the announcements trigger revisions of the forecasts. To compute Δf_1^i under the REE, we solve our model in Section 2 in two cases. We solve the model without any shocks and then re-solve the model, assuming that the first shock is announced. We



(a) Histograms of Δf_1^i

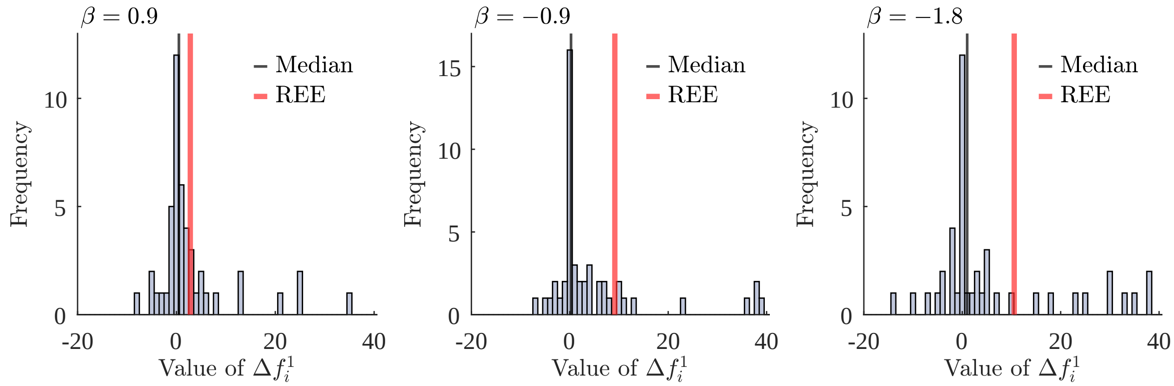
(b) Histograms of Δf_2^i

Figure 4: Forward-Lookingness of Expectations: Case without Announcement

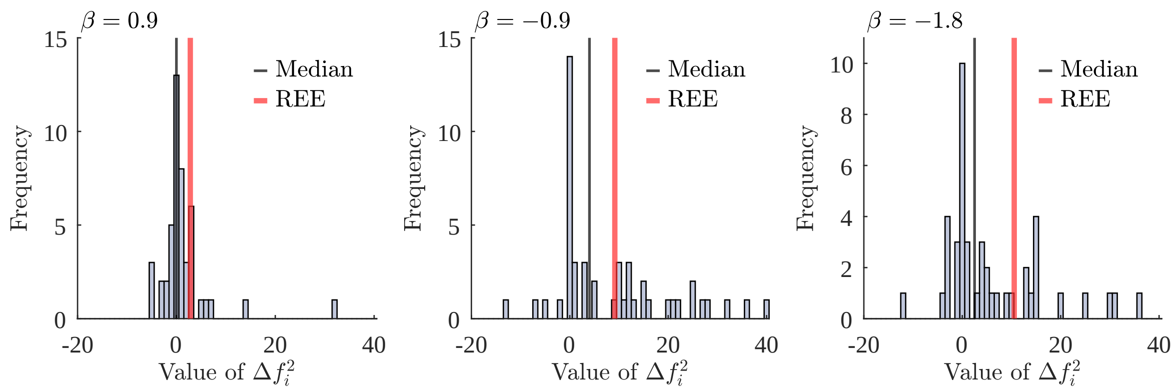
then compute the difference, $P_{14} - P_{13}$, for both cases. Finally, we subtract the difference with no shocks from the difference with the first shock announcement. This double difference corresponds to the theoretical counterpart of Δf_1^i. We compute the theoretical value of Δf_2^i, using the same procedure.

We begin our analysis by considering the scenario where shocks are not announced. This analysis serves as a sanity check for our experimental setup since we naturally expect participants not to revise their forecasts in the absence of new information. Figure 4 presents the histograms for Δf_1^i and Δf_2^i for various β values. The figure indicates that both Δf_1^i and Δf_2^i for all β are centered around zero, with their median values exactly equal to zero.[29] As discussed above, this result aligns with the model's prediction under the REE; most participants do not revise their forecasts unless new information arrives. Nevertheless, it is

[29]The fractions of individuals who do not revise their forecasts are greater than 55%.



(a) Histograms of Δf_1^i



(b) Histograms of Δf_2^i

Figure 5: Forward-Lookingness of Expectations: Case with Announcement

important to note that our finding constitutes a necessary condition for participants to hold rational expectations, but it does not suffice on its own. For example, if participants strongly believe that the environment is stationary and they would assume that $P_t = P_{t+1} = \cdots$, leading to $E^i(P_{t+1} - P_t | P_{\tau<t}) = 0$. In such cases, even though participants do not form their expectations based on rational expectations, Δf_1^i and Δf_2^i still equal zero. We examine how each participant forms her belief over the price levels in Section 5.4.

Next, we analyze the case where shocks are announced, as depicted in the histograms in Figure 5. Several important observations emerge from this analysis. First, the histograms appear right-skewed, and the median values have become positive, indicating that some participants respond significantly to the announcements. Second, there is considerable heterogeneity in participants' responses. To illustrate this heterogeneity more clearly, we compute for each treatment in Table 3 the fraction of participants who did not revise their fore-

First Announcement	Overreaction	Underreaction	
		No Revisions	Other Types
$\beta = 0.9$	27%	17%	56%
$\beta = -0.9$	19%	27%	54%
$\beta = -1.8$	30%	21%	49%

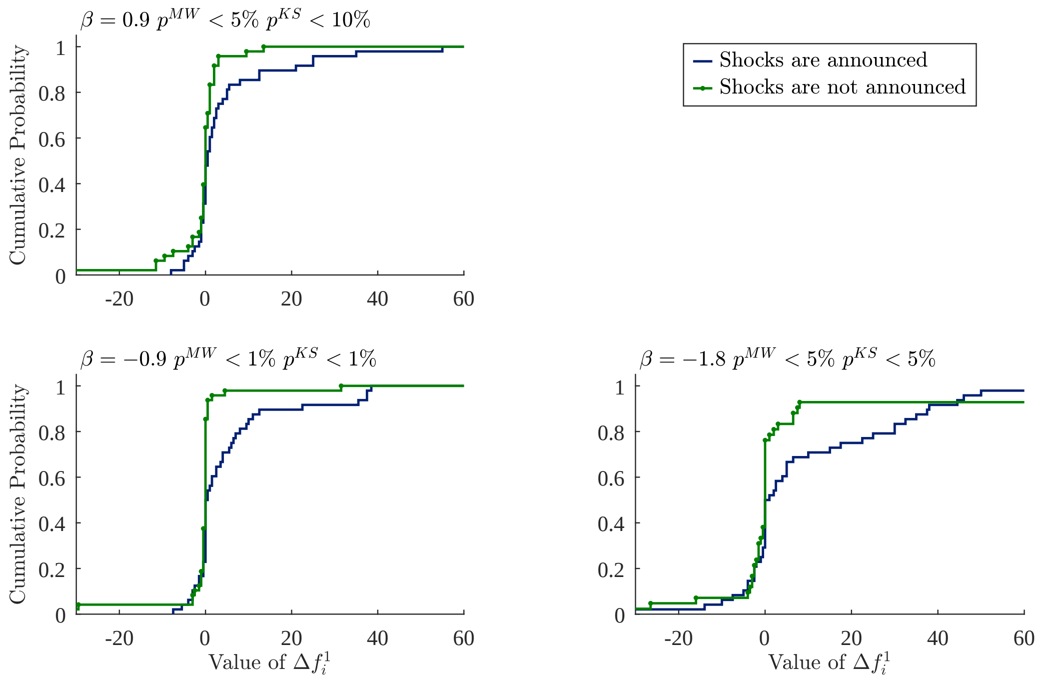
Second Announcement	Overreaction	Underreaction	
		No Revisions	Other Types
$\beta = 0.9$	13%	25%	62%
$\beta = -0.9$	44%	29%	57%
$\beta = -1.8$	29%	19%	57%

Table 3: Heterogeneity of Revisions

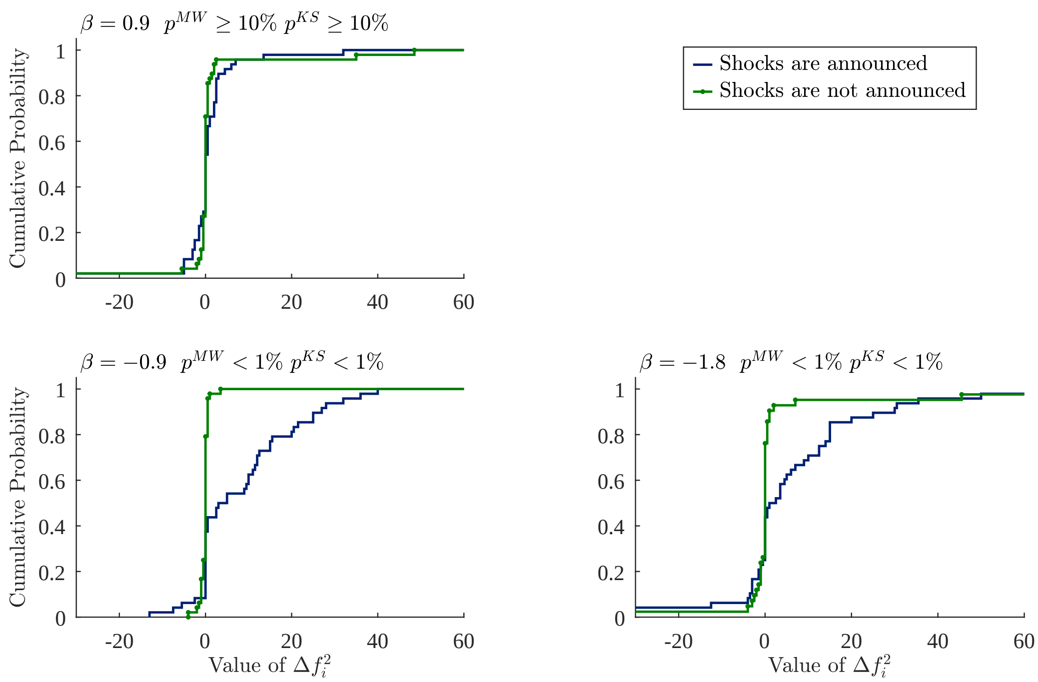
Notes: The Overreaction column shows the proportion of participants who adjusted their forecasts more than the REE model predicts. The Underreaction column is divided into two categories. No Revisions indicates the proportion of participants who did not alter their price forecasts at all, calculated as the percentage of participants for whom $\Delta f_j^i = 0$. Other Types represents the proportion of participants who did not overreact but still changed their price forecasts.

casts ($\Delta f_j^i = 0$) and those who revised their forecasts to a greater extent than predicted by the REE model. As shown in Figure 5, about a quarter of participants did not revise their forecasts, implying that they ignored the announcement when forecasting price change. However, a significant fraction of participants adjust their forecasts upward, accounting for the announcements. Notably, some even overreact by revising their forecasts more than the REE model predicts. For example, after the announcement of the second shock, 40% of participants overreacted when $\beta = -0.9$. This finding suggests strong forward-looking behavior.

Given the considerable heterogeneity in participants' responses, it becomes essential to formally examine whether participants respond to the announcement, on average. To this end, we analyze whether the mean values of $(\Delta f_J^i)_{J=1,2}$ differ between conditions with and without the announcement using the Mann-Whitney U test. Furthermore, we investigate whether these values are drawn from the same distribution by applying the Kolmogorov-Smirnov test. We hypothesize that the announcement causes a rightward shift in the distri-



(a) Empirical Distribution of Δf_1^i



(b) Empirical Distribution of Δf_2^i

Figure 6: Forward-Lookingness of Expectations

Notes: The p-value for the Mann-Whitney U test is denoted by p^{MW}, while the Kolmogorov-Smirnov test is represented by p^{KS}.

bution of Δf^i_J; therefore, we perform these tests with a one-sided alternative hypothesis.

Figure 6 presents the empirical distributions of Δf^i_1 and Δf^i_2 for both announced and unannounced cases, along with the associated p-values from the tests. As shown in Figure 6, the distributions tend to shift to the right, except in the case of $\beta = 0.9$ and $J = 2$. The statistical tests confirm these visual patterns. Specifically, the distribution of Δf^i_J with an announcement is significantly different from that without an announcement, suggesting that participants generally revised their forecasts upward in response to an announcement. We conclude this section by stating that participants responded to the announcements, on average, indicating that they are forward-looking.

5.4 Individual Expectation Formation

In the previous section, we analyzed how participants react to shocks or their announcements *on average*. Unfortunately, understanding how each participant forms her beliefs about the shocks and announcements is challenging since we essentially have only two shocks and the associated announcement (two observations) per participant. Instead of understanding how individuals respond to the shocks and their announcements, we examine how each participant i forms expectations in the absence of shocks and announcements in this section. In other words, we investigate how each participant forms her beliefs in a multi-period environment during "normal" times.

Motivated by the work of Anufriev and Hommes (2012), we estimate a reduced-form forecasting rule. We generalize their forecasting rule by allowing coefficients to vary across different forecast horizons:

$$f^i_{t,t+j} = \alpha^i_j + \gamma^i_j P_{t-1} + \omega^i_j \left(f_{t-1,t-1} - P_{t-1}\right) + \chi^i_j \left(P_{t-1} - P_{t-2}\right) + \varepsilon^i_{t,t+j}. \tag{10}$$

This reduced-form learning equation can capture various expectation formation mechanisms. For example, both adaptive heuristic and trend-following methods can be represented within this framework. Participant i uses an adaptive heuristic if the forecast rule is given by

$$f^i_{t,t+j} = P_{t-1} + \omega^i_j \left(f_{t-1,t-1} - P_{t-1}\right).$$

Here, the coefficient ω^i_j indicates the extent to which participant i incorporates her past expectation into her current expectation. Alternatively, participant i adopts a trend-following method if her expectation is given by

$$f^i_{t,t+j} = P_{t-1} + \chi^i_j \left(P_{t-1} - P_{t-2}\right),$$

where χ^i_j governs the strength of extrapolation; a higher χ^i_j results in greater extrapolation by the participant.

Note that the reduced-form forecasting rule in Eq. (10) is entirely backward-looking. As discussed previously, some participants respond to the shocks and announcements. To concentrate on belief formation during stable periods, we eliminate the influence of forward-looking behavior by excluding data from periods with shocks and announcements, if any. Because we drop these observations, we do not differentiate between cases with and without announcements.

We begin our analysis by running the regression in Eq. (10) for each participant i. Similar to Anufriev and Hommes (2012), the reduced-form forecasting rule in Eq. (10) captures most of the variations in forecasts. Table 4 reports the average R^2 for each treatment. The R^2 values are high across all treatments and horizons, but vary with respect to both. Specifically, the R^2 decreases as the forecast horizon increases. Additionally, when $\beta = -0.9$, the R^2 is higher than in other cases, likely reflecting the fact that prices fluctuate less when $\beta = -0.9$.

Individual-level regressions allow us to categorize participants into different types. Table 5 presents these classifications across various treatments and horizons, based on whether the coefficients from the regression in Eq. (10) are significant at $p < 0.1$. Participants are classified into types represented by triplets (e.g., 1-0-0), where each digit indicates the statistical significance (1) or non-significance (0) of the respective coefficients (γ^j_i, ω^j_i, χ^j_i). The

Treatment	Horizon j				
	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
$\beta=0.9$	0.88	0.85	0.82	0.79	0.77
$\beta=-0.9$	0.91	0.87	0.85	0.84	0.80
$\beta=-1.8$	0.80	0.77	0.74	0.74	0.70

Table 4: Average R^2 by Treatment

(1-0-0) type makes up a large portion of participants, suggesting that many rely exclusively on past price information (P_{t-1}) in their forecasting. In contrast, very few participants do not use past price information at all (0-*-*). Additionally, the (1-1-1) type is also common, indicating that many participants employ both adaptive heuristic and trend-following strategies in their forecasts. Types (1-0-1) and (1-1-0) are also significantly represented, demonstrating the diversity in how participants form their expectations. Thus, Table 5 reveals substantial heterogeneity in expectation formation among participants.

While Table 5 indicates how participants use past information, it does not reveal whether the magnitudes of the coefficients are economically meaningful. To assess the sizes of the coefficients, we report their conditional means and associated standard deviations among significant coefficients in Table 6. Three key observations emerge. First, participants predominantly use the most recent price, P_{t-1}, as a reference point when forming their forecasts, as evidenced by coefficients on P_{t-1} being close to one across all forecast horizons.

Second, participants consider their own past forecast errors, $f_{t-1,t-1} - P_{t-1}$, indicating that individual expectations are somewhat self-referential. While the coefficients could theoretically be negative, the regression results suggest they are positive and close to 0.5. This implies that participants gradually adjust their expectations even if their previous forecast exceeded the previous price, i.e., $f^i_{t-1,t-1} > P_{t-1}$.

Finally, the influence of the most recent price change, $P_{t-1} - P_{t-2}$, becomes more pronounced as the forecast horizon extends, when the game exhibits positive feedback. Specifically, the coefficients χ^i_j on $P_{t-1} - P_{t-2}$ increase substantially from 0.53 for $j = 0$ to 1.95 for

Treatment	Horizon	Type							
		1-0-0	1-1-1	1-1-0	1-0-1	0-1-0	0-1-1	0-0-1	0-0-0
	$j = 0$	29%	38%	15%	15%	0%	1%	0%	3%
	$j = 1$	24%	41%	16%	16%	0%	1%	1%	2%
$\beta = 0.9$	$j = 2$	28%	39%	14%	14%	1%	2%	1%	2%
	$j = 3$	27%	42%	15%	11%	0%	1%	1%	3%
	$j = 4$	30%	35%	14%	14%	0%	4%	1%	2%
	$j = 0$	34%	25%	26%	11%	1%	1%	1%	0%
	$j = 1$	38%	18%	27%	14%	1%	0%	2%	1%
$\beta = -0.9$	$j = 2$	32%	22%	29%	12%	1%	0%	2%	1%
	$j = 3$	36%	23%	24%	12%	1%	0%	2%	1%
	$j = 4$	40%	22%	22%	10%	1%	0%	2%	3%
	$j = 0$	27%	14%	30%	18%	2%	0%	1%	8%
	$j = 1$	30%	16%	33%	13%	0%	0%	0%	8%
$\beta = -1.8$	$j = 2$	29%	19%	27%	14%	3%	1%	0%	7%
	$j = 3$	32%	19%	27%	12%	1%	0%	1%	8%
	$j = 4$	32%	19%	24%	14%	1%	0%	1%	8%

Table 5: Classification of Types Across Treatments and Horizons

Notes: We run separate regressions, $f^i_{t,t+j} = \alpha^i_j + \gamma^i_j P_{t-1} + \omega^i_j(f_{t-1,t-1} - P_{t-1}) + \chi^i_j(P_{t-1} - P_{t-2}) + \varepsilon^i_{t,t+j}$, for each participant and determine the p-values of the coefficients. The three elements in the triplets indicate the significance (1) or non-significance (0) of the coefficients γ^i_j , ω^i_j, and χ^i_j, respectively.

$j = 4$ when $\beta = 0.9$. This significant increase suggests that participants place greater importance on recent price trends when making longer-term forecasts, relying more on momentum or trend-following behavior for expectations about the distant future.

This last point mertis further attention. In typical experimental settings where only one-period expectations are elicited, it is not possible to ascertain whether participants view price increases as persistent or temporary. This point becomes clearer when we compare the coefficients χ^i_0 when $\beta = 0.9$ to those when $\beta = -0.9$. Both coefficients are around 0.5, suggesting that participants anticipate continued price increases based on recent trends. If this were the only evidence available, one would conclude that participants employ similar

Treatment	Coefficients	Horizon j				
		$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
$\beta=0.9$	γ_j^i	1.02 (0.06)	1.04 (0.08)	1.05 (0.10)	1.11 (0.13)	1.19 (0.18)
$\beta=-0.9$	γ_j^i	0.98 (0.05)	0.96 (0.06)	0.95 (0.06)	0.95 (0.07)	0.94 (0.08)
$\beta=-1.8$	γ_j^i	0.96 (0.06)	0.90 (0.08)	0.89 (0.09)	0.97 (0.10)	0.90 (0.11)
$\beta=0.9$	ω_j^i	0.42 (0.11)	0.52 (0.15)	0.58 (0.21)	0.74 (0.25)	0.63 (0.32)
$\beta=-0.9$	ω_j^i	0.34 (0.10)	0.38 (0.13)	0.46 (0.18)	0.50 (0.17)	0.56 (0.19)
$\beta=-1.8$	ω_j^i	0.36 (0.11)	0.36 (0.13)	0.72 (0.34)	0.22 (0.19)	0.28 (0.15)
$\beta=0.9$	χ_j^i	0.53 (0.18)	0.75 (0.25)	1.23 (0.31)	1.74 (0.52)	1.95 (0.46)
$\beta=-0.9$	χ_j^i	0.45 (0.22)	0.45 (0.22)	0.58 (0.26)	0.77 (0.27)	0.90 (0.30)
$\beta=-1.8$	χ_j^i	-0.17 (0.16)	0.03 (0.16)	0.47 (0.42)	-0.03 (0.21)	-0.06 (0.26)

Table 6: Conditional Mean of Coefficients with Standard Errors

Notes; Recall that we run separate regressions, $f^i_{t,t+j} = \alpha^i_j + \gamma^i_j P_{t-1} + \omega^i_j(f_{t-1,t-1} - P_{t-1}) + \chi^i_j(P_{t-1} - P_{t-2}) + \varepsilon^i_{t,t+j}$, for each participant. We compute the average values and standard errors of the coefficients, conditional on their associated p-values being less than 10%.

extrapolation when $\beta = 0.9$ and $\beta = -0.9$. However, our experiment can differentiate whether participants perceive price increases as temporary or persistent. Because multiple forecasts are elicited, we can estimate χ^i_j for $j = 1, \cdots, 4$. As noted above, when $\beta = 0.9$, participants expect higher and higher prices, indicating that they believe the trend will persist. Conversely, when $\beta = -0.9$, participants do not clearly expect prices to keep rising, suggesting they view the price increase as temporary. This highlights the novelty of our experiment in allowing us to discern participants' perceptions of the persistence of price changes.

To further examine this feature, we run a regression with interaction terms

$$f^i_{t,t+j} - f^i_{t,t} = \alpha^i + \mathrm{FE}_j + \gamma^i P_{t-1} + \omega^i \left(f_{t-1,t-1} - P_{t-1}\right) + \chi^i \left(P_{t-1} - P_{t-2}\right)$$
$$+ \tilde{\gamma}^i j P_{t-1} + \tilde{\omega}^i j \left(f_{t-1,t-1} - P_{t-1}\right) + \tilde{\chi}^i j \left(P_{t-1} - P_{t-2}\right),$$

where FE_j represents the fixed effects controlling for horizons. Our primary interest lies in the coefficients $\tilde{\gamma}^i$, $\tilde{\omega}^i$, and $\tilde{\chi}^i$, which govern the strength of the interaction terms. From the

Treatment	Term	Mean Estimate (Std)	$p < 10\%$
$\beta = 0.9$	$\tilde{\gamma}^i$	0.05 (0.01)	60%
$\beta = -0.9$	$\tilde{\gamma}^i$	-0.03 (0.01)	58%
$\beta = -1.8$	$\tilde{\gamma}^i$	-0.02 (0.02)	56%
$\beta = 0.9$	$\tilde{\omega}^i$	0.06 (0.03)	53%
$\beta = -0.9$	$\tilde{\omega}^i$	0.05 (0.02)	47%
$\beta = -1.8$	$\tilde{\omega}^i$	-0.03 (0.01)	38%
$\beta = 0.9$	$\tilde{\chi}^i$	**0.43 (0.07)**	**54%**
$\beta = -0.9$	$\tilde{\chi}^i$	0.11 (0.02)	44%
$\beta = -1.8$	$\tilde{\chi}^i$	0.01 (0.03)	43%

Table 7: Summary of Interaction Terms

Note: We report the conditional means of the coefficients with standard deviations in parentheses, along with the percentage of participants for whom the coefficients are significant at a level below 10% level.

argument above, we would expect $\tilde{\chi}^i$ to be significant and positive.

The regression results are summarized in Table 7. As in previous tables, we report the conditional means of these coefficients among cases where the p-values are less than 10%. The associated standard deviations are reported in parentheses, along with the fraction of participants whose coefficients exhibit statistical significance at levels below 10%.

Table 7 shows that while participants exhibit statistically significant coefficients on the interaction terms involving horizon and other variables, these effects are economically negligible for all cases except those involving the most recent price change, $P_{t-1} - P_{t-2}$. Specifically, consider the interaction terms with P_{t-1} and $(f_{t-1,t-1} - P_{t-1})$. Although these interaction coefficients, $\tilde{\gamma}^i$ and $\tilde{\omega}^i$, are statistically different from zero for a substantial fraction of participants, their magnitudes remain small. In other words, even though some participants adjust their reliance on current price levels or past forecast errors as they look further into the future, these adjustments are tiny. Thus, for these variables, changes in horizon have essentially no meaningful effect on participants' forecasts.

Turning to the interaction involving $P_{t-1} - P_{t-2}$, we first consider cases with strategic substitutes ($\beta < 0$). Here, while the interaction coefficients are often statistically significant, they also remain economically small. In these environments, participants do not meaningfully increase their reliance on recent price trends when forming longer-horizon forecasts, implying that even if prices have risen recently, participants tend to view such increases as temporary and not indicative of persistent inflation.

In contrast, when facing strategic complements ($\beta > 0$), the interaction coefficients for $P_{t-1} - P_{t-2}$ are large enough to be economically important. For example, a 10-unit increase in $P_{t-1} - P_{t-2}$ leads to a 16-unit increase in five-periods-ahead price forecasts under strategic complementarity. Participants extrapolate upward trends more aggressively as the forecast horizon extends and anticipate that rising prices will persist. This new effect of strategic environment – horizon-dependent extrapolation occurring only in environments with strategic complementarity – is a novel finding within our new multi-periods setting.

Our findings have important implications for monetary policy, which delicately depends on how agents form expectations. In environments characterized by strategic complementarity, our results demonstrate that agents' long-run inflation expectation could be de-anchored. This suggests that central banks need to be mindful of these variations in expectation formation, especially because it is costly to re-anchor inflation expectations. Conversely, in environments with strategic substitutes, the nature of forecasting remains relatively stable across horizons. Thus, central banks could conduct their monetary policy without worrying too much about de-anchoring, which simplifies their operations.

6 Concluding Remarks

This paper introduces and experimentally tests a multi-period extension of the standard beauty contest game, inspired by the Calvo (1983) pricing model prevalent in macroeconomics. By eliciting forecasts for multiple future periods, our new design allows us to examine how expectations evolve over time and how agents respond to anticipated changes.

Exploiting this multi-period setup, we propose a novel test for forward-looking behavior. Our results show that participants revise their forecasts after announcements of future shocks, demonstrating that, on average, they incorporate information about future states into their current expectations. This analysis would not be feasible within the constraints of traditional one-period designs, thereby emphasizing the importance of multi-period forecasting experiments in evaluating the forward-lookingness of expectations.

We also uncover significant heterogeneity in individual expectation formation. While many participants rely predominantly on the most recent price level and their past forecast errors, the manner in which they extrapolate trends varies with the strategic environment. Specifically, only in the presence of strategic complementarity do participants strongly extrapolate current price changes into the long run, a pattern that can lead to de-anchored expectations. Under strategic substitution, by contrast, price movements are perceived as more temporary, resulting in more stable long-run expectations.

These findings open several avenues for future research. Of particular interest is the role that variations in the strategic environment play in shaping individual forecasting rules, and how policies can best steer expectations in environments prone to self-fulfilling dynamics. Our multi-period experimental design can serve as a useful tool for examining this issue.

References

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

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 formation and (in)efficient markets in learning-to-forecast and -optimize experiments," *The 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.

Bosch-Domènech, Antoni and Nicolaas J. Vriend, "Imitation of Successful Behaviour in Cournot Markets," *The 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.

Cooper, Kristen, Henry Schneider, and Michael Waldman, "Limited rationality and strategic environment: Further evidence from a pricing game," *Journal of Behavioral and Experimental Economics*, 2021, *90*, 101632.

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

Duffy, John, "Macroeconomics: A survey of laboratory research," in John H. Kagel and Alvin E. Roth, eds., *Handbook of Experimental Economics, Volume 2*, Princeton, NJ: Princeton University Press, 2017, pp. 1–90.

__ **and Daniela Puzzello**, "Experimental Evidence on the Essentiality and Neutrality of Money in a Search Model," *Research in Experimental Economics*, 2014, *17*, 259–311.

__ **and** __ , "The Friedman Rule: Experimental Evidence," *International Economic Review*, May 2022, *63* (2), 671–698.

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," *American Economic Journal: Macroeconomics*, Forthcoming.

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 ed., Princeton, NJ: Princeton University Press, 2015.

García-Schmidt, Mariana and Michael Woodford, "Are Low Interest Rates Deflationary? A Paradox of Perfect-foresight Analysis," *American Economic Review*, 2019, *109* (1), 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 Luba 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 Luba 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, NJ: Princeton University Press, 2003.

A Equivalence

In this section, we begin by introducing the price forecast game used in our experiment. We then show that the (equilibrium outcome of) subgame-perfect Nash equilibrium coincides with the REE of the multi-period beauty contest game discussed in the paper.

Before moving forward, we restate the characterization result for the REE of our multi-period beauty contest game. It follows from Proposition 1 that a price sequence $(P_{t+s})_{s=0}^{\infty}$ is a rational expectation equilibrium, given P_{t-1}, if and only if for all $s \geq 0$,

$$P_{t+s} = (1-\theta) P_{t+s-1} + \theta \sum_{k=0}^{T} \frac{(1-\gamma\theta)^k}{\sum_{l=0}^{T} (1-\gamma\theta)^l} (\alpha + \beta P_{t+s+k}). \tag{A.1}$$

A.1 Setup for the Price Forecast Game

There is a continuum of identical individuals uniformly distributed over $[0,1]$, with time being discrete and indexed by $t = 0, 1, 2, \ldots$. The game terminates with probability γ, which is given exogenously. In each period, individuals submit their forecasts for a sequence of aggregate prices over the next $T+1$ periods. The aggregate prices are endogenously determined by these submitted forecasts. Let $\mathbf{f}_t^i = \left(f_{t,t+s}^i\right)_{s=0}^{T}$ denote individual i's forecast sequence submitted in period t. We assume that each element of $\mathbf{f}_t^i$ belongs to the space $\mathcal{A} = [a, b]$, where $0 < a < b$. The aggregate price P_t is assumed to also take a value from $\mathcal{A}$.

We characterize the environment in terms of histories. We denote H^t as the set of public histories that records the sequence of the aggregate prices up to period t:[30]

$$H^t \equiv \mathcal{A}^t = \mathcal{A} \times \mathcal{A} \times \cdots \times \mathcal{A}.$$

An element of H^t is denoted by h^t, which is a sequence of past public events (aggregate prices):

$$h^t = (P_{-1}, P_0, \cdots, P_{t-1}),$$

[30]Technically, the termination event is part of the public information, but it is unnecessary to track it. The game ends at this point, and there are no subsequent subgames.

where P_{-1} is the given initial condition. Let H denote the set of all histories, $H = \cup_{t \geq 0} H^t$.[31]

The static payoff for individual i in period t is given by

$$\Pi\left(F_t^i - P_t\right) = 100/\left(1 + \left|F_t^i - P_t\right|\right),$$

where F_t^i is a payoff-relevant forecast of individual i, which is defined as follows. As mentioned earlier and in Section 3, individual i can submit her forecasts $\mathbf{f}_t^i$ every period t. With probability $1-\theta$, the submitted forecast determines the payoff-relevant forecast F_t^i : that is, $F_t^i = f_{t,t}^i$. Let ω_t^i denote an indicator that takes the value of one if i's submitted forecasts in period t become payoff-relevant in the same period and zero otherwise. If the submitted forecasts do not become payoff-relevant, then the payoff-relevant forecast in period $t-1$ remains effective in period t, so $F_t^i = f_{s,t}^i$, where s denotes the most recent period before period t when the submitted forecasts became relevant. If the submitted forecasts fail to become payoff-relevant for $T+1$ consecutive periods, then the next forecasts become payoff-relevant with certainty. Moreover, the initial payoff-relevant forecasts $\left(F_{-1}^i\right)_i$ are given.

The mapping from these forecasts, $(\mathbf{f}_t^i)_i$, to the aggregate price in period t is given by

$$P_t = \theta P_{t-1} + \int_{\left\{i;\omega_i^t=1\right\}} \sum_{s=0}^{T} \frac{(1-\gamma\theta)^s}{\sum_{l=0}^{T}(1-\gamma\theta)^l}\left(\alpha + \beta f_{t,t+s}^i\right) di. \tag{A.2}$$

For a micro-foundation for Eq. (A.2), see Section 2.

The present value of individual i's payoff in period t is given by

$$E_t \sum_{s=0}^{\infty} \gamma^s \Pi\left(F_{t+s}^i - P_{t+s}\right), \tag{A.3}$$

The expectation operator is taken only over F_{t+s}^i, because, with probability $1-\theta$, the payoff-relevant forecast may change. Individual i maximizes this objective function by choosing appropriate $(\mathbf{f}_t^i)_{t=0}^{\infty}$.

Formally, the strategy of individual i, σ^i, is a function from H to $\mathcal{A}^{T+1}$. A price function σ^P is a function from H to $\mathcal{A}^\infty$. This price function maps the current history to a sequence

[31] As becomes evident shortly, private histories do not affect the payoff or feasibility. Therefore, we do not need to track them.

of the prices:

$$\sigma^P\left(h^t\right) = \left(P_{t+s}\right)_{s=0}^{\infty}.$$

This sequence $\sigma^P\left(h^t\right)$ is referred to as the continuation outcome path in h^t.

Note that from σ^i and σ^P, the sequence of forecasts $\left(\mathbf{f}_{t+s}^i\right)_{s=0}^{\infty}$ is uniquely determined, which in turn uniquely determines the distribution of $\left(F_{t+s}^i\right)_{s=0}^{\infty}$. Let $\left(\mathbf{f}_{t+s}^i\left(h^t;\sigma^i,\sigma^P\right)\right)_{s=0}^{\infty}$ denote the forecasts induced from h^t and $\left(F_{t+s}^i\left(h^t;\sigma^i,\sigma^P\right)\right)_{s=0}^{\infty}$ the induced payoff-relevant forecasts, which are random variables. An equilibrium requirement imposes a cross restriction on σ^P and $\left(\sigma^i\right)_{i\in[0,1]}$.

We now formally define a *subgame-perfect Nash equilibrium* in this dynamic game.

Definition 1. A subgame-perfect Nash equilibrium in this large game is a profile of the strategies $\left(\sigma^i\right)_{i\in[0,1]}$ and a price function σ^P such that (i) for all history $h^t \in H$, $\left(\mathbf{f}_{t+s}^i\left(h^t;\sigma^i,\sigma^p\right)\right)_{s=0}^{\infty}$ maximizes the expected present-value payoff in Eqs. (A.3), and (ii) the consistency condition is satisfied. For all history $h^t \in H$,

$$P_{t+s} = \theta P_{t+s-1} + \int_{\omega_t^i=1} \sum_{k=0}^{T} \frac{(1-\gamma\theta)^k}{\sum_{l=0}^{T}(1-\gamma\theta)^l}\left(\alpha + \beta f_{t+s,t+s+k}^i\right) di \quad \text{for all } s \geq 0, \tag{A.4}$$

where $\left(P_{t+s}\right)_{s=0}^{\infty} = \sigma^P\left(h^t\right)$.

A.2 Equivalence Result

Before proving the equivalence result, we characterize the strategy profiles.

Lemma 1. *In a subgame-perfect Nash equilibrium, for any $h^t \in H$,*

$$\sigma^i\left(h^t\right) = \left(P_{t+s}\right)_{s=0}^{T}, \tag{A.5}$$

where $\left(P_{t+s}\right)_{s=0}^{T}$ corresponds to the first $T+1$ elements of $\sigma^P\left(h^t\right)$.

Proof. If $\sigma^i\left(h^t\right) = \left(P_{t+s}\right)_{s=0}^{T}$ for all h^t, then $F_{t+s}^i\left(h^t;\sigma^i,\sigma^p\right) = P_{t+s}$ for all $s \geq 0$. Thus, the static payoff in period $t+s$ is maximized. If Eq. (A.5) is violated, there is a future event such that Π is not maximized, and this event happens with positive probability. Thus, in any subgame-perfect Nash equilibrium, Eq. (A.5) follows. $\square$

An immediate corollary of Lemma 1 is that individuals choose the same strategy, $\sigma^i = \sigma$.

We now proceed to establish the equivalence result.

Proposition 2. *Fix any history h^t such that $h_t = P_{t-1}$. The sequence of the aggregate prices, $(P_{t+s})_{s=0}^{\infty}$, is the rational expectation equilibrium price given P_{t-1} if and only if $(P_{t+s})_{s=0}^{\infty}$ is the continuation outcome path of the subgame perfect equilibrium of $\left((\sigma^i)_i, \sigma^P\right)$ given $h_t = P_{t-1}$.*

Proof. In any subgame-perfect equilibrium, the optimal strategy for individuals is given by Eq. (A.5) and $\sigma^i = \sigma$. Thus, the consistency condition in Eq. (A.4) reduces to the following equation: for all $s \geq 0$,

$$P_{t+s} = \theta P_{t+s-1} + \theta \sum_{k=0}^{T} \frac{(1-\gamma\theta)^k}{\sum_{l=0}^{T} (1-\gamma\theta)^l} \left(\alpha + \beta P_{t+s+k}\right). \tag{A.6}$$

The continuation outcome of the subgame-perfect Nash equilibrium given any $h_t = P_{t-1}$ satisfies Eq. (A.6), which is equivalent to Eq. (A.1). Thus, $(P_{t+s})_{s=0}^{\infty}$ is the rational expectation equilibrium price given P_{t-1} if and only if the sequence is the continuation outcome of the subgame-perfect Nash equilibrium given any $h_t = P_{t-1}$. $\square$

Our results, Proposition 2, and Lemma 1, have two implications for our experimental design. First, Proposition 2 validates the use of the learning-to-forecast game over the original dynamic beauty contest game. A similar result was first established by Marimon and Sunder (1993). Second, from Lemma 1, individuals maximize their payoffs by minimizing forecast errors. This finding applies to games with strategic substitutes or complements and forms the basis for our experimental design.

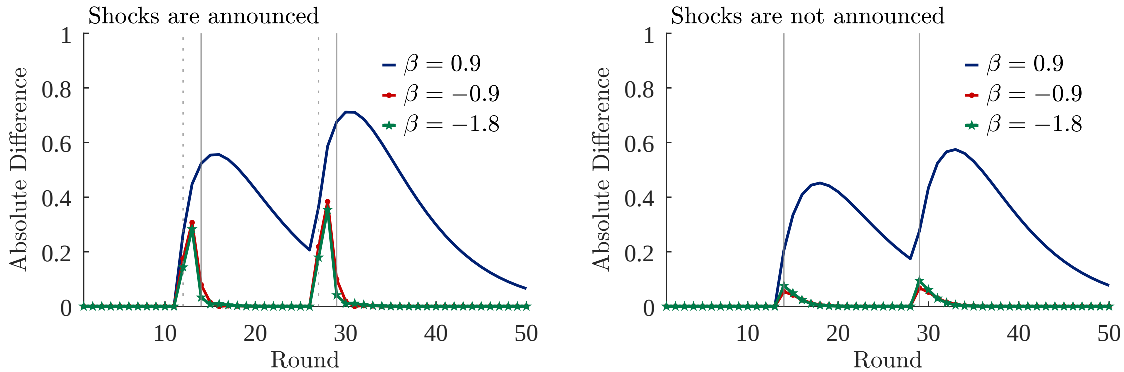


Figure A.1: Comparison between the REE Prices with and without Termination Probability

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

In our experiments, we inadvertently ignored the impacts of termination probability, γ, 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 noted in footnote 12. 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^γ, P_t^γ) denote the equilibrium individual price and the equilibrium price when $\gamma \geq 0$. Figure A.1 depicts the absolute difference between $P_t^{5\%}$ and P_t^0 in cases with and without announcement. Note that the differences are less than one in all the periods, including those around the shocks.

C Detail of the Procedure in 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 his or her microphone is turned on only when necessary.

Upon connecting to a Zoom session, participants first remain 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 a given participant ID in the form of "sub##", where ## is the two-digit number that is valid during the experiment. Once participants are assigned an ID, 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 if their internet connection fails during the experiment). Then, the prerecorded instruction video is played. Although participants are not given a hard copy of the instruction slides, they are informed that they can go through that 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 cannot communicate with each other via Zoom chat.

D English Translations of the Instructions and Examples of Screenshots

English translation of the instruction slides can be found at https://osf.io/spgh5.

Figures A.2 to A.5 show examples of the decision screen participants faced in our experiment.

Note: The values of α and β are shown in red and in a large point size in period 1 and when they change (in periods 14 and 29) in all the treatments (see Figure A.4). In other periods, they are shown in black with a regular point size (see Figure A.5).

Figure A.2: 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 α and β are shown in black text below the announcement.

Figure A.3: Screen with a pre-announcement of future shock (only for treatments with pre-announcement)

Note: The values of α and β are shown in red and a large point size when they change (in periods 14 and 29) in all treatments.

Figure A.4: Screen when the shock is realized (common to all treatments)

Note: The values of α and β are shown in black and a regular point size.

Figure A.5: Screen for normal periods (common to all treatments)

D.1 Comprehension quiz

Here is an English translation of the 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. Let's also 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 for only 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.