Experimental Economics (2025), **28**, 1214–1239
doi:10.1017/eec.2026.10048

ORIGINAL PAPER

An experiment on a multi-period beauty contest game

Nobuyuki Hanaki[1,2] and Yuta Takahashi[1]

[1]Institute of Social and Economic Research, the University of Osaka, and University of Limassol, 6-1 Mihogaoka, Ibaraki, Osaka, Japan
[2]University of Limassol, Cyprus
Corresponding author: Nobuyuki Hanaki; Email: nobuyuki.hanaki@iser.osaka-u.ac.jp

(Received 25 January 2024; revised 23 December 2024; accepted 21 March 2026)

Abstract
We present and conduct a novel experiment on a multi-period beauty contest game. Leveraging the multi-period feature, we investigate how participants revise their forecasts in periods when new information—such as shocks or announcements—arrives and how they form their expectations in the absence of new information. We make two key contributions. First, we develop a new method based on forecast revisions to evaluate whether participants behave in a forward-looking manner. The experimental results show that participants do react to anticipated shocks: namely, the announcements of future shocks. Second, we identify a new strategic environment effect during periods without new information; only when the game exhibits strategic complementarity do participants use extrapolation and expect continuously rising prices. This finding suggests that expectation formation is endogenous to the economic environment; and policy design should thus take this endogeneity into account.

Keywords: Expectation formation; forward-lookingness; learning-to-forecast experiment; strategic environment effect
JEL Code: C92; D84; E70

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 and have proposed alternative frameworks (e.g., Angeletos & Lian, 2018; Farhi & Werning, 2019; García-Schmidt & Woodford, 2019).

Experimental evidence from "learning-to-forecast" experiments further calls into question 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 in the presence of positive expectational feedback between the aggregate outcome and individual forecasts (Bao et al., 2017; Heemeijer et al., 2009; Hommes et al., 2005).[1] The presence of *expectational feedback*—where

[1]Note that unlike experiments in which participants actually trade in the market (e.g., Akiyama et al., 2017; Asparouhova et al., 2016; Crockett et al., 2019; Haruvy et al., 2007; Smith et al., 1988), set prices (e.g., Fehr & Tyran, 2008; Noussair et al., 2015; Orland & Roos, 2013; Orland & Roos, 2019; Petersen, 2015), or set quantities (e.g., Bosch-Domènech & Vriend, 2003; Huck et al., 1999; Huck et al., 2004; Offerman et al., 2002), participants in learning-to-forecast experiments do not trade, set

© The Author(s), 2026. Published by Cambridge University Press on behalf of the Economic Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (http://creativecommons.org/licenses/by-nc-sa/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

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 & Hommes, 2012; Evans et al., 2025).[2]

However, most existing learning-to-forecast experiments are essentially of the single-period type.[3] Participants only need to forecast the outcome of today or in one future period (e.g., tomorrow), and the realized outcome is a function of submitted forecasts regarding this single period (seeBao et al., 2021, for a survey).[4] 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 the 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.

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. Based on this new experimental framework, we examine how individuals revise their expectations in response to new information, as well as how they form forecasts during normal periods without such information. We aim to contribute to the literature by providing this new dynamic experimental framework, offering a new method to assess the forward-lookingness of expectations, and identifying a new strategic environmental effect.

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

prices, or determine quantities; they only forecast. This eliminates the need for participants to trade or set prices (or quantities) 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 those authors call "learning-to-optimize" experiments).

[2]Learning-to-forecast experiments can be broadly categorized into those based on univariate environments and those based on multivariate environments. The former correspond mostly to those based on an asset-pricing model, such as Hommes et al. (2005), Heemeijer et al. (2009), Bao et al. (2017), Anufriev et al. (2022a), and Evans et al. (2022), as well as simple demand and supply dynamics, as in Bao and Duffy (2016), Bao and Hommes (2019), and Evans et al. (2025). The last paper, Evans et al. (2025), is one of the most closely related with our paper. The main difference is that we study a multi-period beauty contest game and analyze how participants respond to anticipated shocks, not unanticipated shocks. Experiments using multivariate environments largely correspond to those based on the New Keynesian macroeconomic model, such as Adam (2007), Pfajfar and Zakelj (2014); Pfajfar and Zakelj (2018), Arifovic and Petersen (2017), Hommes et al. (2019), Rholes and Petersen (2021), Assenza et al. (2021), Mauersberger (2021), Petersen and Rhoes (2022), Ahrens et al. (2023), Lustenhouwer and Salle (2025), as well as those based on more abstract settings (Anufriev et al., 2022b; Levelt et al., 2021). These multivariate learning-to-forecast experiments based on the New Keynesian model are used to investigate policy-relevant questions such as the effectiveness of different monetary policy rules (Assenza et al., 2021; Cornand & M'Baye, 2018; Kostyshyna et al., 2022; Mauersberger, 2021; Pfajfar & Zakelj, 2014; Pfajfar & Zakelj, 2018), provisions of central bank projections and communications (Ahrens et al., 2023; Arifovic et al., 2023; Kryvtsov & Petersen, 2021; Mokhtarzadeh & Petersen, 2021), and the effectiveness of these policies at the zero lower bound (Arifovic & Petersen, 2017; Hommes et al., 2019).

[3]There are a few recent exceptions in both univariate and multivariate environments. In the univariate environment, Colassante et al. (2020) elicit a series of forecasts for multiple future periods; however, unlike our framework, the forecasts beyond the one for the next period do not determine the market outcome. In the multivariate environment, Adam (2007), Rholes and Petersen (2021), and Petersen and Rhoes (2022) elicit one- and two-period-ahead forecasts of inflation rates, and determine current outcomes. In Lustenhouwer and Salle (2025), participants submit their forecasts for output in multiple future periods, and these forecasts determine the current outcome. Our proposed experiment is in the univariate environment, and all the submitted forecasts can potentially determine the current outcome.

[4]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).

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

reduces the demand for that firm, demand may increase or decrease when other firms raise their prices. If firms can set their prices freely, 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.

Following Calvo (1983), we introduce a friction that limits firms' ability to choose the 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 expectational 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 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. We set up our game so that participants merely need to anticipate aggregate prices accurately and their payoffs are maximized when they submit accurate forecasts in each period, regardless of whether the game features strategic substitutes or complements. We formally show the equivalence between this transformed experiment and the original price-setting game in Appendix A.

We adopt a three-by-two between-subjects experimental design that varies two key dimensions. The first dimension pertains to the structure of strategic interaction among participants, contrasting conditions of strong strategic complementarity with those involving equally strong or greater degrees of substitution. The second dimension concerns the timing of information about upward shifts in the demand curve, which we call shocks. In some treatments, participants face shocks that occur unexpectedly. Such shocks are referred to as unanticipated shocks. In others, they are announced two periods in advance. Such *announcements* of future shocks are termed anticipated shocks. This experimental 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 strategic environment effect, which posits that prices deviate substantially from REE prices—that is, prices expected under REE—when the game features strategic complements, whereas prices converge quickly to REE prices otherwise. Re-examination of this well-established strategic environment effect in the literature serves as a sanity check to ensure that our experimental framework produces results that are aligned with those of existing one-period learning-to-forecast games (see, e.g.,Bao et al., 2012; Bao et al., 2017; Bao & Hommes, 2019; Heemeijer et al., 2009).[5] We find that the strategic environment effect is replicated in our new experimental framework, thereby validating the internal consistency of our setup before we move on to more substantive analyses.

Having reproduced the strategic environment effect in our experimental framework, we proceed to investigate the forward-looking nature of expectation formation. In particular, we examine how participants respond to anticipated shocks; namely, the announcement of future shocks. Recall that the 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 for periods after the realization of the shock following the announcement by comparing the

[5]The strategic environment effect has also been demonstrated in other contexts, including a price-setting game (Cooper et al., 2021; Fehr & Tyran, 2008; Funaki et al., 2023), a duopoly game (Potters & Suetens, 2009), and one-shot beauty contest games (Hanaki et al., 2019; Sutan & Willinger, 2009). More recently, Evans et al. (2022) and Anufriev et al. (2022a) show that the strategic environment effect attenuates in dynamic settings that differ from ours. Their frameworks are based on asset-pricing models and elicit forecasts either for the average across multiple future periods (Evans et al., 2022) or for a single distant period (Anufriev et al., 2022a).

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

forecasts for the same future periods before and after the announcement. 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 shocks have materialized.

Our experimental data reveal significant heterogeneity in how participants respond to the anticipated shocks. While certain individuals forecast price levels that surpass those implied by the REE, others completely ignore anticipated shocks. The latter response is consistent with level-0 reasoning. This result is similar to the one based on the responses to the unanticipated shocks reported by Evans et al. (2025) in their single-period setting.

Although the heterogeneity in behavior is noteworthy in its own right, our experimental framework permits us to direct attention to the average responses of participants and to evaluate whether these are consistent with forward-looking behavior.[6] Our analysis demonstrates that, on average, participants do respond to the anticipated shocks in line with forward-looking behavior. This finding provides new evidence of forward-lookingness in expectation formation, offering a distinctive contribution to the literature.

After examining participants' responses to anticipated shocks, we turn to the formation of expectations during periods that lack such shocks. That is, we focus on "normal" times to study belief formation in a dynamic environment. Following Anufriev and Hommes (2012), we estimate a reduced-form forecasting rule that encompasses various heuristics like adaptive expectations and trend following. We generalize their approach by allowing coefficients of these heuristics 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 impact of recent price *changes* becomes stronger as the forecast horizon lengthens. This effect is particularly strong in positive feedback environments, where participants increasingly rely on trend-following behavior. This finding suggests that in positive feedback environments (i.e., those with strategic complementarity), participants expect price increases to persist, potentially leading to de-anchored long-run inflation expectations. Conversely, in negative feedback settings (i.e., those with strategic substitutability), 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 operates to effectively conduct monetary policy and mitigate the risks of expectation de-anchoring.

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 and derive the equilibrium price under the REE of the model. The theoretical model serves as the foundation for our experiment design, and the REE price offers a benchmark for comparison in the subsequent analysis.

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

[6] In our experimental framework, only the average forecast is of primary importance. In broader contexts, however, it may be necessary to examine higher-order moments of the forecast distribution—a line of inquiry reserved for future research.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

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

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

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

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).

We now formally define an *REE* as a pair of prices (p_t, P_t) such that given the aggregate price $(P_t)_t$, p_t solves the maximization problem (Eq. 3), and 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*

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

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 Eq. (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 the fraction $1-\theta$ of firms can reset their prices. Because firms that cannot reset their prices today are randomly chosen, the average price among them is P_{t-1}. Firms that can reset their prices choose the same price level given by Eq. (4), so the average price P_t satisfies Eq. (5).*

[7] In New Keynesian models, the time horizon T is typically set to infinity, consistent with the assumption of firms with infinite lives. By contrast, experimental constraints make it unfeasible to collect an unbounded sequence of forecasts. Accordingly, we impose a finite time horizon T in our experimental design.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

Notably, $\alpha + \beta P_t$ represents the optimal price in period t if 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 optimal flexible prices. Eq. (4) captures the dynamic thought process within firms; the optimal price, p_t, depends on current actions by others, and on 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 & 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 explicitly requiring participants to anticipate not only others' actions in period t but also in subsequent periods $t + 1, \dots, t + T - 1$.

3. Experimental design

In this section, we describe the experimental setup by detailing the structure of the learning-to-forecast game that participants actually play. We then outline our treatment design, followed by an explanation of the online experimental procedure.

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 participants' rewards, as detailed below, all five forecasts have the potential to do so when submitted.[10]

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

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

where P_t is the realized price in period t, and F^i_t is the payoff-relevant forecast of participant i in period t, as defined below.[11] 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^i_t is set to $F^i_1 = f^i_{1,1}$. In period 2, the payoff-relevant forecast F^i_2 is determined probabilistically: with probability $1 - \theta$, the newly submitted set of forecasts become payoff-relevant, yielding $F^i_2 = f^i_{2,2}$; with probability θ, the previous forecasts remain payoff-relevant, resulting in $F^i_2 = f^i_{1,2}$. 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 introduced in Section 2. The probability of firms being given an opportunity to re-optimize their prices in

[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 close to 100 participants (Hommes et al., 2021).

[9] All forecasts are constrained to be integers.

[10] Our experimental design may prompt participants to adopt a more forward-looking reasoning compared to existing univariate learning-to-forecast experiments because we ask participants to provide forecasts for several future periods rather than only one.

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

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

Table 1. Hypothetical submitted forecasts

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

period t, $1-\theta$, is translated into the probability of the new set of forecasts becoming payoff-relevant for participants. The horizon, T, over which firms optimize 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 prices for periods 1 to 5:

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

In period 1, this set of forecasts becomes payoff-relevant so that the payoff-relevant forecast is $F^i_1 = 10$. In period 2, the participant submits new forecasts $f^i_{2,2}, \cdots, f^i_{2,6}$. Suppose that the new forecasts do not become payoff-relevant in period 2. Then, $F^i_2 = f^i_{1,2} = 11$. If the new set of forecasts becomes payoff-relevant in periods 3 and 4, then the latest forecasts determine the payoff-relevant forecasts. Thus, $F^i_3 = f^i_{3,3} = 9$ and $F^i_4 = f^i_{4,4} = 13$.[12] The reward for the participant is determined by the difference between the payoff-relevant forecasts F^i_t and the aggregate price P_t.

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

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

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 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 G for the screenshots.

In our experiment, we set the reset probability θ to $1/2$ and assume that the game ends with probability $\gamma = 0.05$ at the end of each period.[13] Participants are rewarded based on the total points they earn throughout the game. As noted above, we set $T = 5$. Eq. (6) can be used to provide a 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

[12] If the new forecasts in period 3 and 4 do not become payoff-relevant, then $F^i_3 = f^i_{1,3} = 12$ and $F^i_4 = f^i_{1,4} = 12$.

[13] 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 (see Appendix F). As shown in Appendix B, however, this oversight results in the REE price differing only slightly (by less than one). Furthermore, because P_t is rounded to the nearest integer, we believe this oversight does not affect the experimental outcomes.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

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 0.027. 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, either 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 & 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.[14] 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 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.

To minimize variation in experiment duration and participant payments across sessions, we followed the procedure of Duffy and Puzzello (2014); Duffy and Puzzello (2022) by predefining the random number sequence. Predefining ensured that all sessions repeated two blocks and concluded after the first game. Consistent with Duffy and Puzzello (2014); Duffy and Puzzello (2022), this information was not disclosed to participants.[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 at period 29. We assume that the initial flexible REE price is 65; 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 varying two main aspects of the games. The first is the degree of strategic interaction. We consider games where β takes values in $\{0.9, -0.9, -1.8\}$; that is, the games exhibit either strategic complementarity (positive feedback) or substitution (negative feedback).[16] We also explicitly consider a case of strong substitution with $\beta = -1.8$. This consideration is motivated by the fact that although New Keynesian models may exhibit strong substitutability, as demonstrated by García-Schmidt and Woodford (2019), the issue has not been investigated to a greater extent in the experimental literature, with exceptions such as Bao and Duffy (2016) and Evans et al. (2025).

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

[14] 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.

[15] In the instructions, participants are informed 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. If the game is repeated multiple times, one of the games will be randomly selected for payment; see Appendix F.

[16] Because participants are restricted to submitting integer-valued forecasts, multiple equilibria may emerge under strong strategic complementarities. However, all such equilibria are local to the REE examined in the main text. As a result, this multiplicity does not materially affect our findings. See Appendix D for a detailed discussion.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

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 two periods before its realization, that is, 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 realization of shocks*, both before and after the announcements.[17] Because we elicit participants' forecasts only for five future periods, if the announcement is made either too early in advance, it becomes impossible to analyze its immediate impact on forecast revisions at the time the shock occurs. Therefore, we make the announcement two periods prior to the occurrence of the shock and analyze how forecasts are revised, the question that lies at the core of our study.

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, and until they finished the comprehension quiz. To begin the first game, all participants had to correctly answer all six quiz questions.[19] Final rewards were provided through Amazon Gift Cards (e-mail version).

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

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 in 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 in advance, $\alpha_{t,k} = \alpha_{t,t}$ for all k and t. When the shocks occur, $\alpha_{t,t}$ suddenly increases. Suppose that the shocks are pre-announced. In that case, $\alpha_{t,k}$ changes when the shocks are announced, not when they occur. Again, the pre-announcement of the future shocks, anticipated 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 anticipated shocks have 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 they are not announced (right panel). These figures can be intuitively understood. When the game demonstrates strategic complementarity ($\beta > 0$), the transition to new

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

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

[19] We did not add a button for participants to review the instructions on the main decision screen during the experiment because we considered having correctly answered all the questions in the comprehension quiz to demonstrate that they understood the instructions.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

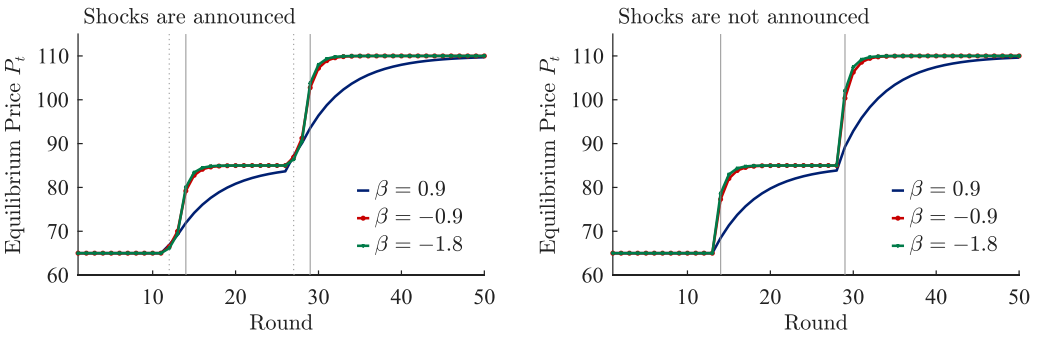


Fig 1. Aggregate price P_t under rational expectation
Note: The solid vertical lines indicate when the two shocks occur, while the dotted lines in the left panel show when these shocks are announced.

steady state equilibrium prices is slower due to Calvo pricing friction. Participants recognize that others may not adjust their prices swiftly because of that 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 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. Analysis of the experimental data

In this section, we begin by outlining the experimental sessions, including participant numbers, treatment assignments, and earnings. Then, we proceed to analyze our experimental data.

5.1. Overview of the experimental sessions

We conducted our experiments in April and May 2023, involving a total of 294 participants.[20] 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.[21] 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 $\beta = 0.9$, respectively.

5.2. 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

[20] However, in one session with an 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.

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

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

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.

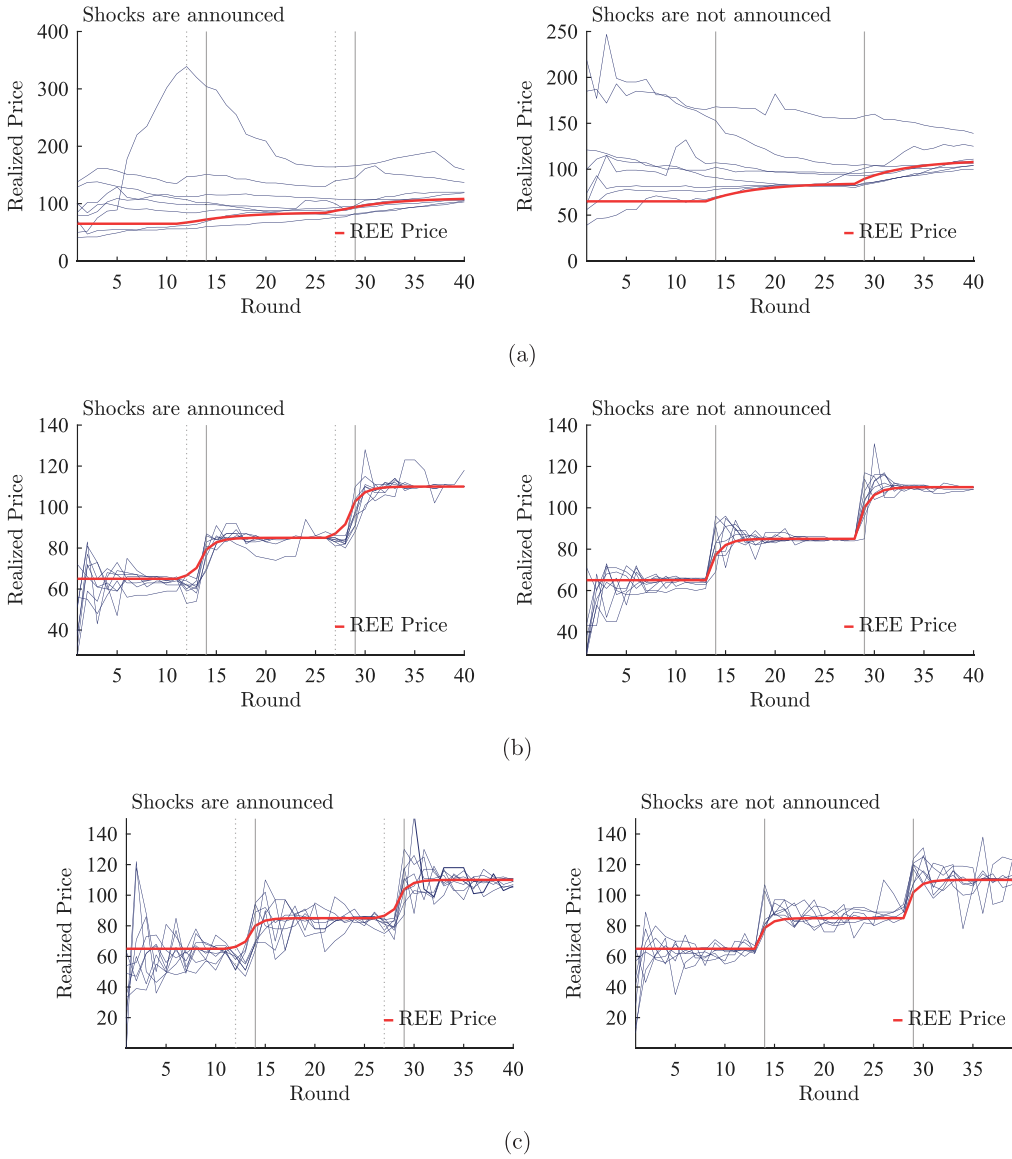


Fig 2. Realized aggregate prices P_t (a) Positive Feedback: $\beta = 0.9$ (b) Weak Negative Feedback: $\beta = -0.9$ (c) Strong Negative Feedback: $\beta = -1.8$

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 vertical lines in the left panel show when these shocks are announced.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

prices follow the REE price closely when the game exhibits strategic substitutes $\beta < 0$. When the game exhibits strategic complements $\beta > 0$, they deviate persistently from the REE prices. As shown by Fehr and Tyran (2008), these features can be understood intuitively. When the game exhibits strategic complements, 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, and 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, this mechanism does not operate. Participants want to choose higher (lower) prices when others choose lower (higher) prices. Thus, their initial expectations are not self-fulfilling unless they coincide with the REE price, and they often quickly converge toward 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), Bao et al. (2012); Bao et al. (2017), and Bao and Hommes (2019).

It is evident that aggregate prices become more stable when the game has weak strategic substitutes ($\beta = -0.9$). By contrast, with strong strategic substitutes ($\beta = -1.8$), aggregate prices hover around the REE price but exhibit greater deviations. A similar pattern is also documented by Bao and Duffy (2016) and Evans et al. (2025). This result likely stems from difficulties in expectation coordination. When β becomes more negative, the price depends sensitively on expectations. This sensitivity results in the observed instability of prices.[22]

5.3. 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, P_t^{REE} is the REE price in period t, and K represents the total number of periods (40 except for the two groups that faced a technical problem).[23]

Figure 3 shows the empirical cumulative distributions of RADs (top) and RDs (bottom) in the treatments with (left) and without (right) an announcement. In each panel, the distributions for each β are shown. The top panels show that regardless of the existence of an 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.[24] Furthermore, the distributions of RADs are ranked

[22] This result seems to contradict Assenza et al. (2021), Pfajfar and Zakelj (2014); Pfajfar and Zakelj (2018), and Mauersberger (2021), who study the effectiveness of monetary policy in the New Keynesian learning-to-forecast experiment. These studies demonstrate that a more forceful monetary policy response to deviations from the inflation target enhances the strength of the negative feedback mechanism, thereby improving the stabilization of aggregate dynamics. However, it is possible that even the strongest monetary policy reaction considered in these experiments is not strong enough to make the degree of strategic substitutability sufficiently negative, as considered in our experiment and those by Bao and Duffy (2016) and Evans et al. (2025). This is because the Calvo pricing friction attenuates the impact of monetary policy on the sensitivity of expectation.

[23] 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 13.

[24] They are all significantly different from zero at the 5% level, according to the signed-rank test. The p-values are 0.016 for $\beta = -1.8$ without an announcement and 0.008 for the other five treatments.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

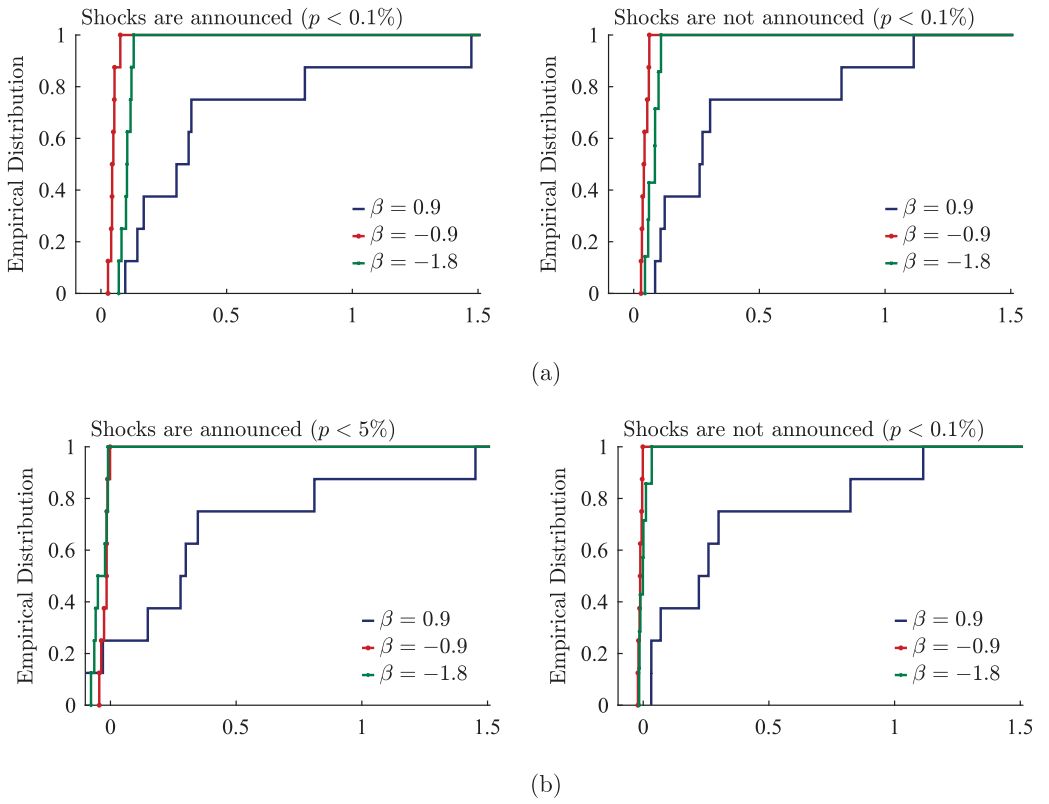


Fig 3. Deviations from the rational expectation prices (a) Relative Absolute Deviation (RAD) (b) Relative Deviation (RD)
Note: *p*-values of Kruskal-Wallis test are reported.

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 $\beta = 0.9$ than with other treatments, and the differences between the three treatments are statistically significant both with and without an announcement.[25] The 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 in Figure 3 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 $\beta = -1.8$.[26] The lack of statistical significance 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 $\beta = -1.8$.[27]

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

[26]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 an 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 an announcement.

[27]In fact, RDs are not significantly different from zero for $\beta = -1.8$ without an 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 an announcement and 0.0078 for the remaining four treatments, signed-rank test).

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

5.4. Forward-lookingness of expectations

We now examine whether forecasts have responded to anticipated shocks: that is, the announcement of future shocks.[28] To accomplish this objective, we analyze forecast revisions before and after anticipated shocks to assess whether individuals have reacted to the shocks or disregarded them. In particular, if there are no revisions to forecasts after an anticipated shock, it suggests that the anticipated shock is ignored.[29]

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 price changes $P_{14} - P_{13}$ and $P_{29} - P_{28}$ in response to the announcements in period 12 and 27, respectively.[30] The conditional mean is calculated based on information available through periods 13 and 28 instead of periods 12 and 27, to mitigate 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 anticipated shocks. In particular, observing nonzero revisions strongly suggests that participants do indeed 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.[31] However, it is important to emphasize that identifying the forward-lookingness through these revisions is a sufficient but not strictly 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 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

[28] We also examine how individuals respond to unanticipated shocks, as analyzed by Evans et al. (2025), in Appendix C.

[29] 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.

[30] 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.

[31] 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.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

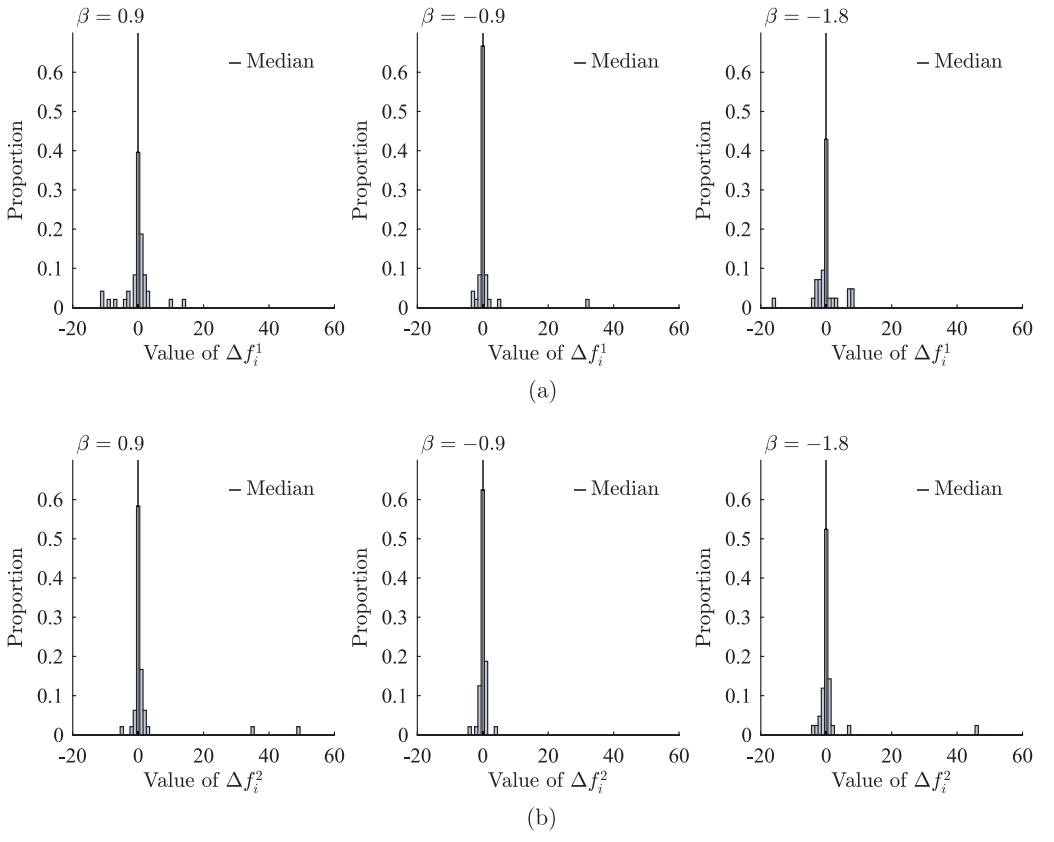


Fig 4. Forward-lookingness of expectations: case without announcement (a) Distributions of Forecast Revisions After the First Announcement Δf_1^i (b) Distributions of Forecast Revisions After the Second Announcement Δf_2^i

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, and both Δf_1^i and Δf_2^i are zero in this case.

Consider the case where shocks are announced. Since those announcements constitute new information, they 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 solve it again under the assumption that the first shock is announced. We 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. To be specific, the fractions of individuals who do

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

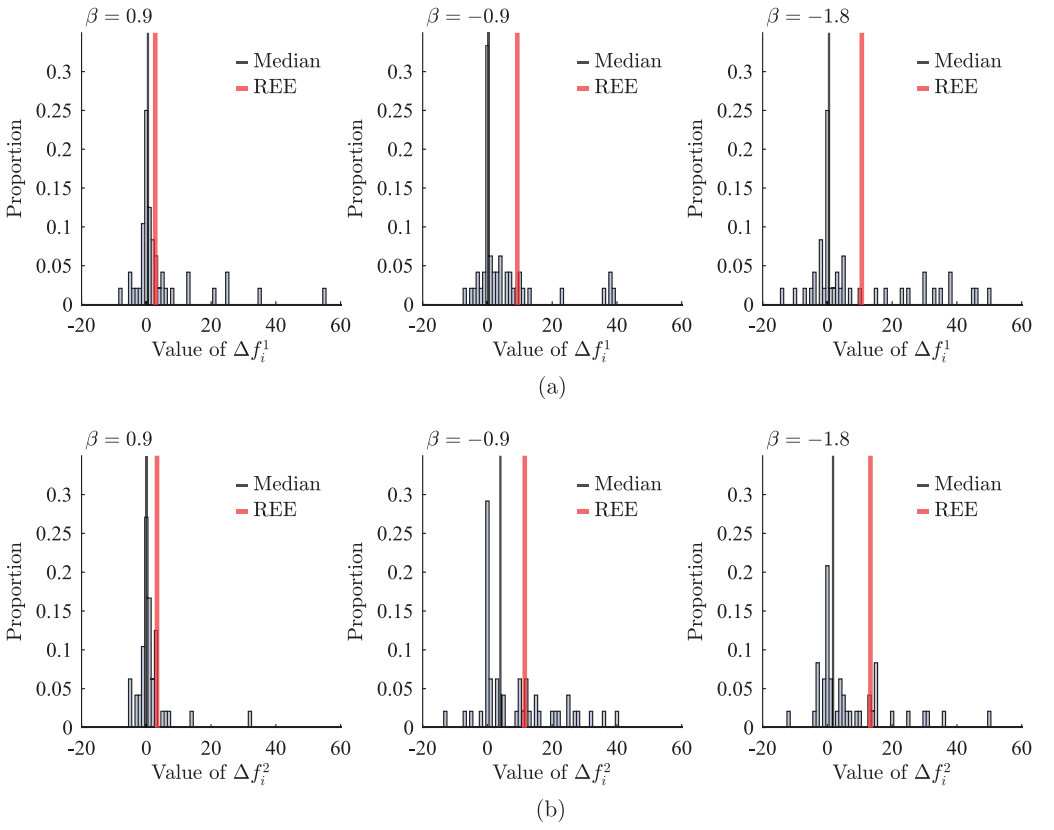


Fig 5. Forward-lookingness of expectations: case with announcement (a) Distributions of Forecast Revisions After the First Announcement Δf_1^i (b) Distributions of Forecast Revisions After the Second Announcement Δf_2^i

not revise their forecasts are greater than 40%. 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. It is important to note that while our finding constitutes a necessary condition for participants to hold rational expectations, it does not suffice on its own. For example, if participants strongly believe that the environment is stationary and would assume that $P_t = P_{t+1} = \cdots$, that leads 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 5.5.

Next, we analyze the case where shocks are announced, as depicted in the histograms in Figure 5. Three 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 examine this heterogeneity, we compute for each treatment the fraction of participants who did not revise their forecasts ($\Delta f_j^i = 0$), those who revised their forecasts to a greater extent than predicted by the REE model, and others. 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

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

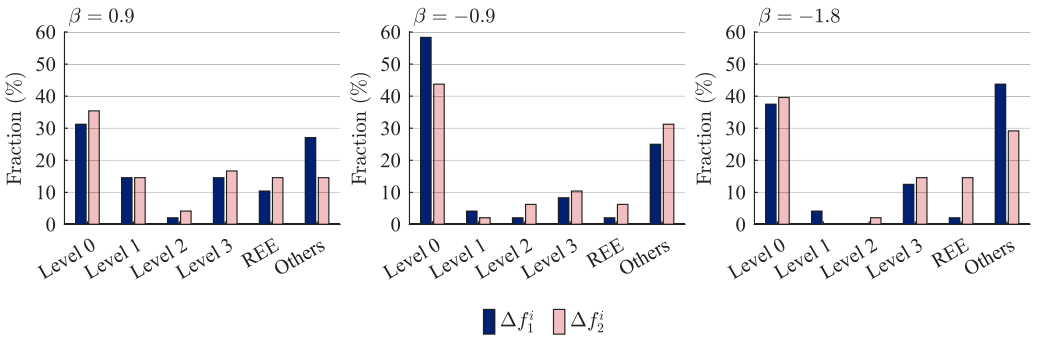


Fig 6. Classification of participants based on depth of reasoning

announcement of the second shock, 33% of participants overreacted when $\beta = -0.9$. This finding suggests strong forward-looking behavior.[32]

Another way to classify revisions, Δf_j^i, is to apply level-k reasoning (Nagel, 1995). The level-k reasoning represents a form of forward-looking expectation formation and is often employed to analyze situations in which agents must make forecasts after unprecedented shocks (Farhi & Werning, 2019). Following Evans et al. (2025), we derive the theoretical counterparts of Δf_j^i under level-k forecasting.[33] Let the default forecast sequence be $P_{t+s}(0)_{s=0}^{\infty}$. The level-$k$ forecast sequence $P_{t+s}(k)_{s=0}^{\infty}$ is then defined recursively by

$$P_t(k) = (1-\theta)\sum_{s=0}^{T-1} \frac{((1-\gamma)\theta)^s}{\sum_{m=0}^{T-1}((1-\gamma)\theta)^m}(\alpha + \beta P_{t+s}(k-1)) + \theta P_{t-1}(k-1). \tag{10}$$

We define the default forecast sequence $P_{t+s}(0)$ as the average of the realized prices in the two periods immediately preceding the forecast date. For example, when computing the level-k theoretical forecast

$$E^i[P_{14} - P_{13} \mid t \leq 13],$$

we set

$$P_{t+s}(0) = \frac{P_{12} + P_{13}}{2} \quad \text{for all } s \geq 0.$$

We then assign each participant to the level whose theoretical forecast lies closest to her observed revision. Participants whose revisions deviate by more than three units from all level-0, level-1, level-2, and REE forecasts are categorized as "Others." The cutoff of three follows Evans et al. (2025). Figure 6 shows the results.

As expected from Figure 5, large shares of participants are classified as level-0 forecasters since they do not revise their forecasts at all. Moreover, sizable shares fall into each of the level-0 through level-3 and REE categories even in our setting. These results mirror those reported by Evans et al. (2025). One notable quantitative divergence is that the share of participants whose forecasts conform to the REE benchmark is smaller here than in Evans et al. (2025). In our setup, only about 5% and

[32]While Figure 5 indicates some heterogeneity across treatments, most differences are not statistically significant. We apply pairwise proportion tests to compare the shares of overreactions, underreactions, and no revisions. In most cases, the p-values exceed 10%. A notable exception is the comparison between $\beta = 0.9$ and $\beta = -0.9$, where the fraction of participants who overreact to the second anticipated shock differs significantly.

[33]Note that our classification exercise differs from that of Evans et al. (2025). They classify the forecasts right after the shocks without pre-announcements, not revisions, and classify current forecasts according to level-k. In Appendix C, we conduct the same exercise conducted by Evans et al. (2025) in our dynamic setting based on the treatment without an announcement.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

12% of participants are classified as REE forecasters after the first and second announcements, respectively, compared with 25% and 38% in Evans et al. (2025). This discrepancy likely reflects the greater difficulty of constructing REE forecasts in our dynamic environment, where agents must account for the transition path to the new steady state.[34]

Finally, we investigate whether participants begin to forecast the REE price more accurately following the announcements. We do so by computing the absolute difference between Δf_j^i and the REE benchmark for each participant and treatment. We then evaluate whether the median of these participant-level deviations declines significantly after the second announcement within each treatment group. Although Figure 5 indicates the possibility of learning in some treatments (notably for $\beta = 0.9$), the observed deviations are not statistically significantly smaller for the second announcement than the first one. Accordingly, we fail to find evidence of learning in response to announcements.[35]

While accounting for the substantial heterogeneity in participants' responses is important, it is equally critical to assess whether, on average, participants react to the announcement. This is relevant because in our linear demand system, the aggregate price is governed by the average forecast. 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 distribution of Δf_J^i; therefore, we perform these tests with a one-sided alternative hypothesis.

Figure 7 presents the empirical distributions of Δf_1^i and Δf_2^i for both announced and unannounced cases, along with the associated p-values from the tests. As shown in that figure, 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_J^i 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, on average, participants responded to the announcements, indicating that they are forward looking.

5.5. Individual expectation formation

Having analyzed average responses to anticipated shocks in the preceding section, we now turn to expectation formation in the absence of shocks. We investigate how each participant forms expectations during such normal periods within our multi-period framework.

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:

$$f_{t,t+j}^i = \alpha_j^i + \gamma_j^i P_{t-1} + \omega_j^i \left(f_{t-1,t-1} - P_{t-1}\right) + \chi_j^i \left(P_{t-1} - P_{t-2}\right) + \varepsilon_{t,t+j}^i. \tag{11}$$

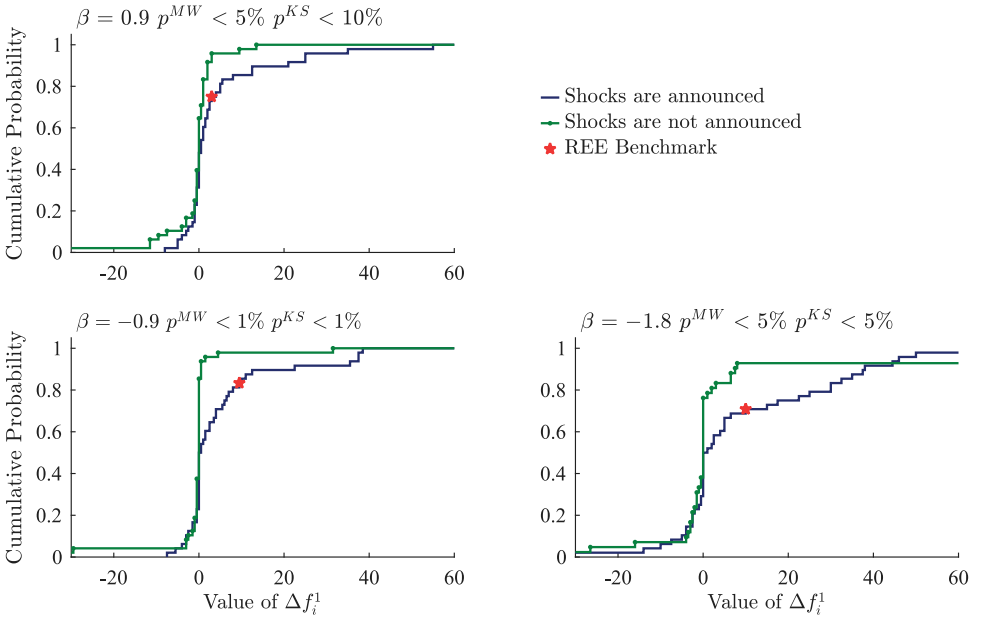
This reduced-form learning equation can capture various expectation formation mechanisms. For example, both adaptive heuristic and trend-following methods can be represented within this framework. Participant i uses an adaptive heuristic if the forecast rule is given by

$$f_{t,t+j}^i = \gamma_j^i P_{t-1} + \omega_j^i \left(f_{t-1,t-1} - P_{t-1}\right).$$

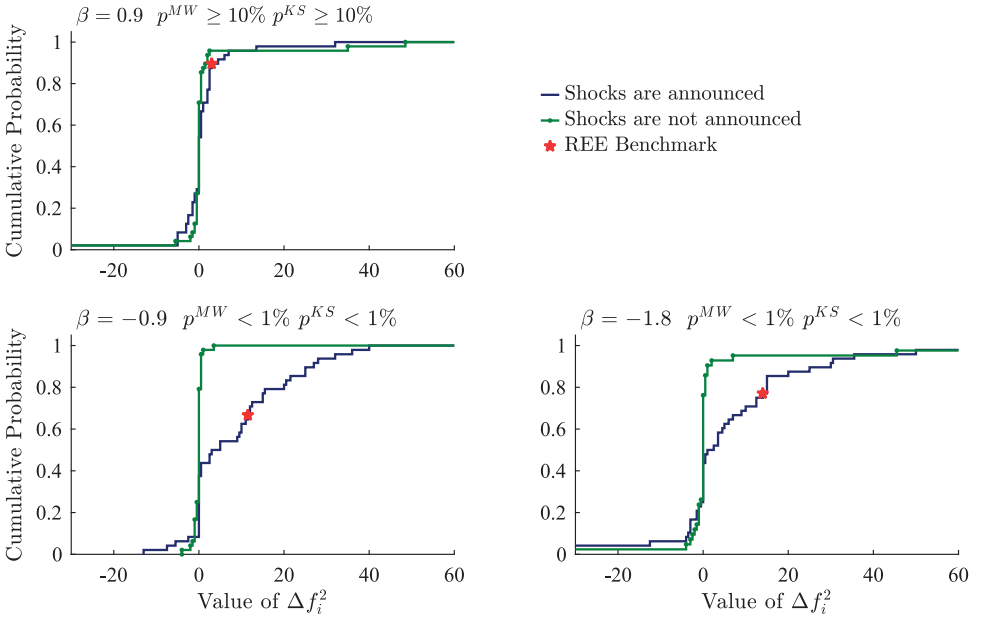
[34] One potential difficulty with our analysis is that unlike Evans et al. (2025), we must specify the sequence of default forecasts. In a dynamic environment such as ours, there is little reason to assume that these defaults remain constant. Exploring how to model level-k forecasts without fixed defaults and how participants select their initial forecast sequence in a dynamic environment is left for future research.

[35] Our experimental design includes only two shocks and two corresponding announcements. If participants were exposed to more than two announcements, it is possible that learning would emerge over time. We leave this question for future research.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press



(a)



(b)

Fig 7. Forward-lookingness of expectations (a) Empirical Distribution of Forecast Revisions After the First Announcement Δf_1^i (b) Empirical Distribution of Forecast Revisions After the Second Announcement Δf_2^i

Notes: The *p*-value for the Mann-Whitney U test is denoted by p^{MW}, while the one for the Kolmogorov-Smirnov test is represented by p^{KS}. The REE benchmarks represent the theoretical values of the revisions, $(\Delta f_j^i)_{j=1,2}$, when shocks are announced. The theoretical values of the revisions when shocks are not announced are trivially zero.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

Table 3. Average R^2 and classification of types across treatments and forecast horizons

			Type							
Treatment	Horizon	R^2	1-0-0	1-1-1	1-1-0	1-0-1	0-1-0	0-1-1	0-0-1	0-0-0
$\beta = 0.9$	$j = 0$	0.88	29%	38%	15%	15%	0%	1%	0%	3%
	$j = 1$	0.85	24%	41%	16%	16%	0%	1%	1%	2%
	$j = 2$	0.82	28%	39%	14%	14%	1%	2%	1%	2%
	$j = 3$	0.79	27%	42%	15%	11%	0%	1%	1%	3%
	$j = 4$	0.77	30%	35%	14%	14%	0%	4%	1%	2%
$\beta = -0.9$	$j = 0$	0.91	34%	25%	26%	11%	1%	1%	1%	0%
	$j = 1$	0.87	38%	18%	27%	14%	1%	0%	2%	1%
	$j = 2$	0.85	32%	22%	29%	12%	1%	0%	2%	1%
	$j = 3$	0.84	36%	23%	24%	12%	1%	0%	2%	1%
	$j = 4$	0.80	40%	22%	22%	10%	1%	0%	2%	3%
$\beta = -1.8$	$j = 0$	0.80	27%	14%	30%	18%	2%	0%	1%	8%
	$j = 1$	0.77	30%	16%	33%	13%	0%	0%	0%	8%
	$j = 2$	0.74	29%	19%	27%	14%	3%	1%	0%	7%
	$j = 3$	0.74	32%	19%	27%	12%	1%	0%	1%	8%
	$j = 4$	0.70	32%	19%	24%	14%	1%	0%	1%	8%

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.

The coefficient ω^i_j captures the extent to which participant i incorporates her past expectations into her current expectation. We generalize the adaptive heuristic in Anufriev and Hommes (2012) by allowing the coefficient on P_{t-1} to differ from one. Alternatively, participant i adopts a trend-following method if her expectation is given by

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

where χ^i_j governs the strength of extrapolation; a higher χ^i_j results in greater extrapolation by the participant. As in the adaptive heuristic, we do not impose a unit coefficient on P_{t-1}.

Note that the reduced-form forecasting rule in Eq. (11) 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 any data from periods with shocks and announcements. 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. (11) for each participant i. Similar to Anufriev and Hommes (2012), the reduced-form forecasting rule in Eq. (11) captures most of the variations in forecasts. Table 3 reports in the column labeled R^2 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, R^2 decreases as the forecast horizon increases. Additionally, when $\beta = -0.9$, 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 3 presents these classifications across various treatments and horizons, based on whether the coefficients from the regression in Eq. (11) are significant at $p < 0.1$. Participants are classified into types represented

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

Table 4. Conditional means and standard deviations of coefficients

		Horizon j				
Treatment	Coefficients	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
$\beta=0.9$	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)

Notes: We run separate regressions, $f_{t,t+j}^i=\alpha_j^i+\gamma_j^i P_{t-1}+\omega_j^i(f_{t-1,t-1}-P_{t-1})+\chi_j^i(P_{t-1}-P_{t-2})+\varepsilon_{t,t+j}^i$, for each participant i. We compute the average values and standard deviations of the coefficients, conditional on their associated p-values being less than 10%.

by triplets (e.g., 1-0-0), where each digit indicates the statistical significance (1) or non-significance (0) of the respective coefficients (γ_i^j, ω_i^j, χ_i^j). The (1-0-0) type, which is characterized by adaptive expectation formation, makes up a large portion of participants, suggesting that many rely exclusively on past price information (P_{t-1}) in their forecasting. By contrast, very few participants do not use past price information at all (0-*-*). Additionally, the (1-1-1) type is common, indicating that many participants employ both adaptive heuristic and trend-following strategies in their forecasts. Types (1-0-1) and (1-1-0), recognized as trend-chasers and adaptive-heuristic forecasters, are also significantly represented, demonstrating the diversity in how participants form their expectations. Thus, Table 3 reveals substantial heterogeneity in expectation formation among participants.

While Table 3 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 4. 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 the coefficients γ_j^i 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 ω_j^i could theoretically be negative, the regression results suggest that 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: that is, $f_{t-1,t-1}^i > P_{t-1}$.

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_j^i on $P_{t-1}-P_{t-2}$ increase substantially from 0.53 for $j=0$ to 1.95 for $j=4$ when $\beta=0.9$. A formal statistical test supports this observed pattern. Specifically, we conduct a one-sided Mann-Whitney test to examine whether the median coefficient rises with the length of the horizon. The p-value for the difference between $j=0$ and $j=1$ is about 8%, dropping to roughly 0.2% for the comparison between $j=0$ and $j=4$. These findings suggest a statistically significant increase in the median coefficient as the horizon extends. 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 merits 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

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

Table 5. Summary statistics of interaction terms by treatment

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%

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 below the 10% level.

persistent or temporary. This point becomes clearer when we compare the coefficients χ_0^i 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 extrapolation when $\beta = 0.9$ and $\beta = -0.9$. However, our experiment can determine whether participants perceive price increases as temporary or persistent. Because multiple forecasts are elicited, we can estimate χ_j^i 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 for each participant i:

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

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. These coefficients capture the change in anchoring, adaptiveness, and trend-chasing as the horizon increases by one period. Based on the discussion above, we would expect $\tilde{\chi}^i$ to be significant and positive when $\beta = 0.9$.

The regression results are summarized in Table 5. As in previous tables, we report the conditional means of these coefficients among cases where the p-values are under 10%. The associated standard deviations are reported in parentheses, along with the fraction of participants whose coefficients exhibit statistical significance below the 10% level.

Table 5 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 $\tilde{\chi}^i$ are often statistically significant, they also remain economically small. In these environments, participants do not meaningfully increase their reliance

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

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-period-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 the 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 depends delicately 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.

Before closing this section, it is worth highlighting a point that warrants caution. As noted in footnote 10, our design may nudge some participants toward forward-looking behavior by explicitly asking them to forecast several future periods rather than only the next period. It is not evident whether individuals would form such expectations in more typical settings without such instructions. One way to explore this possibility is to conduct an experiment in which participants directly choose their optimal price in a dynamic environment, rather than engage in a learning-to-forecast experiment. While this alternative design allows us to analyze how participants form their forecasts without explicitly asking them to forecast several future periods, it may introduce optimization errors by participants. A carefully designed experiment may help isolate expectation formation in a dynamic environment without either optimization errors or conflating it with the nudge effect. We believe that it is important to explore this question since our findings have important policy implications for possible de-anchoring of expectations. We leave such analysis for future research.

6. Concluding remarks

This paper introduces and experimentally examines a multi-period extension of the standard beauty contest game, inspired by the Calvo (1983) pricing model prevalent in macroeconomics. Based on this new dynamic experimental framework, we investigate how participants respond to shocks and announcements and how they form their expectations without such disturbances.

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.

We also identify a new strategic environment effect in individual expectation formation. We find that trend extrapolation varies with the strategic setting. In the presence of strategic complementarity, participants strongly extrapolate current price changes to the long run. Moreover, this trend-chasing behavior intensifies as the forecast horizon lengthens, implying that long-run expectations can become de-anchored, which in turn generates greater macroeconomic instability. Under strategic substitution, by contrast, price movements are perceived as more temporary, and trend extrapolation does not intensify with the forecast horizon, resulting in more stable long-run expectations.

These results have policy implications. In real-world macroeconomic and financial systems, strategic complementarity is prevalent and agents plan over extended horizons. It is often found in the

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

literature that longer planning horizons tend to stabilize aggregate dynamics by allowing agents to anchor on fundamentals. Our results suggest the opposite: in the presence of strategic complementarity, longer horizons can amplify destabilizing momentum-driven dynamics. Policymakers should be aware of this mechanism and recognize that managing expectations requires particular vigilance in such environments. In all cases, policy design should reflect the strategic context in which agents form their expectations.

Finally, our study suggests directions for future work. 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.

Acknowledgements. 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 discussions. 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, 23H00055, and 25H00388 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. The authors used ChatGPT (OpenAI) and Claude (Anthropic) for English language editing and proofreading. All content was reviewed and approved by the authors, who take full responsibility for the accuracy and integrity of the manuscript. The replication material for the study is available at https://doi.org/10.17605/OSF.IO/GECV8.

References

Adam, K. (2007). Experimental evidence on the persistence of output and inflation. *Economic Journal*, *117*(520), 117.

Ahrens, S., Lustenhouwer, J., & Tettamanzi, M. (2023). The stabilizing effects of publishing strategic central bank projections. *Macroeconomic Dynamics*, *27*(3), 826–868.

Akiyama, E., Hanaki, N., & Ishikawa, R. (2017). It is not just confusion! Strategic uncertainty in an experimental asset market. *Economic Journal*, *127* (605), F563–F580.

Angeletos, G. M., & Lian, C. (2018). Forward guidance without common knowledge. *American Economic Review*, *108*(9), 2477–2512.

Anufriev, M., Chernulich, A., & Tuinstra, J. (2022a). Asset price volatility and investment horizons: An experimental investigation. *Journal of Economic Behavior and Organization*, *193*, 19–48.

Anufriev, M., Duffy, J., & Panchenko, V. (2022b). Learning in two-dimensional beauty contest games: Theory and experimental evidence. *Journal of Economic Theory*, *201* (C), 105417.

Anufriev, M., & Hommes, C. (2012). Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *American Economics Journal: Microeconomics*, *4* (4), 35–64.

Arifovic, J., & Petersen, L. (2017). Stabilizing expectations at the zero lower bound: Experimental evidence. *Journal of Economic Dynamics and Control*, *82* (C), 21–43.

Arifovic, J., Salle, I., & Truong, H. (2023) History-dependent monetary policy: A lab experiment and a HENK model. Working paper 4586602, SSRN.

Asparouhova, E., Bossaerts, P., Roy, N., & Zame, W. (2016). Lucas' in the laboratory. *Journal of Finance*, *71*(6), 2727–2779.

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

Bao, T., & Duffy, J. (2016). Adaptive versus eductive learning: Theory and evidence. *European Economic Review*, *83* (C), 64–89.

Bao, T., Duffy, J., & Hommes, C. (2013). Learning, forecasting, and optimizing: An experimental study. *European Economic Review*, *61* (C), 186–204.

Bao, T., Hennequin, M., Hommes, C., & Massaro, D. (2020). Coordination on bubbles in large-group asset pricing experiments. *Journal of Economic Dynamics and Control*, *110* (C), 103702.

Bao, T., & Hommes, C. (2019). When speculators meet suppliers: Positive versus negative feedback in experimental housing markets. *Journal of Economic Dynamics and Control*, *107* (C), 103730.

Bao, T., Hommes, C., & Makarewicz, T. (2017). Bubble formation and (in)efficient markets in learning-to-forecast and -optimize experiments. *The Economic Journal*, *127* (605), F581–F609.

Bao, T., Hommes, C., & Pei, J. (2021). Expectation formation in finance and macroeconomics: A review of new experimental evidence. *Journal of Behavioral and Experimental Finance*, *32* (C), 100591.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

Bao, T., Hommes, C., Sonnemans, J., & Tuinstra, J. (2012). Individual expectations, limited rationality and aggregate outcomes. *Journal of Economic Dynamics and Control*, *36* (8), 1101–1120.

Bosch-Domènech, A., & Vriend, N. J. (2003). Imitation of successful behaviour in cournot markets. *The Economic Journal*, *113* (487), 495–524.

Calvo, G. A. (1983). Staggered prices in a utility maximizing framework. *Journal of Monetary Economics*, *12*(3), 383–398.

Chen, D. L., Schonger, M., & Wicken, C. (2016). oTree - An open-source platform for laboratory, online and field experiments. *Journal of Behavioral and Experimental Finance*, *9* (C), 88–97.

Coibion, O., Gorodnichenko, Y., & Saten, K. (2018). How do firms form their expectations? New survey evidence. *American Economic Review*, *108* (9), 2671–2713.

Colassante, A., Alfarano, S., Camacho-Cuena, E., & Gallegati, M. (2020). Long-run expectations in a learning-to-forecast experiment: A simulation approach. *Journal of Evolutionary Economics*, *30* (1), 75–116.

Cooper, K., Schneider, H., & Waldman, M. (2021). Limited rationality and strategic environment: Further evidence from a pricing game. *Journal of Behavioral and Experimental Economics*, *90* (C), 101632.

Cornand, C., & M'Baye, C. K. (2018). Does inflation targeting matter? An experimental investigation. *Macroeconomic Dynamics*, *22* (2), 362–401.

Crockett, S., Duffy, J., & Izhakian, Y. (2019). An experimental test of the Lucas asset pricing model. *Review of Economic Studies*, *86* (2), 627–667.

Duffy, J. (2017). Macroeconomics: A survey of laboratory research. In J. H. Kagel and A. E. Roth (Eds.) *Handbook of experimental economics* (Vol 2, pp.1–90) Princeton University Press.

Duffy, J., & Puzzello, D. (2014). Experimental evidence on the essentiality and neutrality of money in a search model. *Research in Experimental Economics*, *17*, 259–311.

Duffy, J., & Puzzello, D. (2022). The friedman rule: Experimental evidence. *International Economic Review*, *63*(2), 671–698.

Evans, G. W., Gibbs, C. G., & McGough, B. (2025). A unified model of learning to forecast. *American Economic Journal: Macroeconomics*, *17* (2), 101–33.

Evans, G. W., Hommes, C., McGough, B., & Salle, I. (2022). Are long-horizon expectations (De-)stabilizing? Theory and experiments. *Journal of Monetary Economics*, *132* (C), 44–63.

Farhi, E., & Werning, I. (2019). Monetary policy, bounded rationality, and incomplete market. *American Economic Review*, *109* (11), 3887–3928.

Fehr, E., & Tyran, J. -R. (2008). Limited rationality and strategic interaction: The impact of the strategic environment on nominal inertia. *Econometrica*, *76* (2), 353–394.

Fréchette, G. R., & Yuksel, S. (2017). Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination. *Experimental Economics*, *20* (2), 279–308.

Funaki, Y., Kawamura, K., Uto, N., & Ueda, K. (2023) Price stickiness and strategic uncertainty: An experimental study, CAMA Working Paper 63/2023. Crawford School of Public Policy, Australian National University.

Galí, J. (2015). *Monetary policy, Inflation and the Business cycle*. (2nd ed.) Princeton University Press.

García-Schmidt, M., & Woodford, M. (2019). Are low interest rates deflationary? A paradox of perfect-foresight analysis. *American Economic Review*, *109* (1), 86–120.

Greiner, B. (2015). An online recruitment system for economic experiments. *Journal of the Economic Science Association*, *1* (1), 114–125.

Hanaki, N., Koriyama, Y., Sutan, A., & Willanger, M. (2019). The strategic environment effect in beauty contest games. *Games and Economic Behavior*, *113* (C), 587–610.

Haruvy, E., Lahav, Y., & Noussair, C. N. (2007). Traders' expectations in asset markets: experimental evidence. *American Economics Review*, *97* (5), 1901–1920.

Heemeijer, P., Hommes, C., Sonnemans, J., & Tuinstra, J. (2009). Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. *Journal of Economic Dynamics and Control*, *33* (5), 1052–1072.

Hommes, C., Kopányi-Peuker, A., & Sonnemans, J. (2021). Bobbles, crashes and information contagion in large-group asset market experiments. *Experimental Economics*, *24* (2), 414–433.

Hommes, C., Massaro, D., & Salle, I. (2019). Monetary and fiscal policy design at the zero lower bound: Evidence from the lab. *Economic Inquiry*, *57* (2), 1120–1140.

Hommes, C., Sonnemans, J., Tuinstra, J., & van de Velden, H. (2005). Coordination of expectations in asset pricing experiments. *Review of Financial Studies*, *18* (3), 955–980.

Huck, S., Normann, H. -T., & Oechssler, J. (1999). Learning in cournot oligopoly - An experiment. *Economic Journal*, *109* (454), C80–C95.

Huck, S., Normann, H. -T., & Oechssler, J. (2004). Two are few and four are many: Number effects in experimental oligopolies. *Journal of Economic Behavior and Organization*, *53* (4), 435–446.

Kostyshyna, O., Petersen, L., & Yang, J. (2022). A horse race of monetary policy regimes: An experimental investigation. Working Paper 30530, National Bureau of Economic Research.

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press

Kryvtsov, O., & Petersen, L. (2021). Central bank communication that works: Lessons from lab experiments. *Journal of Monetary Economics*, *117* (C), 760–780.
Levelt, E., Hommes, C. H., & Hennequin, M. (2021) Planar learning to forecast market games. Working paper 3882461, SSRN.
Lustenhouwer, J., & Salle, I. (2025). Learning to be rational in the presense of news: A lab investigation. *European Economic Review*, *172* (C), 104948.
Marimon, R., & Sunder, S. (1993). Indeterminacy of equilibria in a hyperinflationary world: Experimental evidence. *Econometrica*, *61* (5), 1073–1107.
Mauersberger, F. (2021). Monetary policy rules in a non-rational world: A macroeconomic experiment. *Journal of Economic Theory*, *197* (C), 105203.
Mokhtarzadeh, F., & Petersen, L. (2021). Coordinating expectations through central bank projections. *Experimental Economics*, *24* (3), 883–918.
Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economics Review*, *85* (5), 1313–1326.
Noussair, C. N., Pfajfar, D., & Zsiros, J. (2015). Pricing decisions in an experimental dynamic stochastic general equilibrium economy. *Journal of Economic Behavior and Organization*, *109* (C), 188–205.
Offerman, T., Potters, J., & Sonnemans, J. (2002). Imitation and belief learning in an oligopoly experiment. *Review of Economic Studies*, *69* (4), 973–997.
Orland, A., & Roos, M. W. M. (2013). The New Keynesian Phillips curve with myopic agents. *Journal of Economic Dynamics and Control*, *37* (10), 2270–2286.
Orland, A., & Roos, M. W. M. (2019). Price-setting with quadratic adjustment costs: Experimental evidence. *Journal of Economic Behavior and Organization*, *163* (C), 88–116.
Petersen, L. (2015). Do expectations and decisions respond to monetary policy?. *Journal of Economic Studies*, *42* (6), 972–1004.
Petersen, L., & Rhoes, R. (2022). Macroeconomic expectations, central bank communication, and background uncertainty: A COVID-19 laboratory experiment. *Journal of Economic Dynamics and Control*, *143* (C), 104460.
Pfajfar, D., & Zakelj, B. (2014). Experimental evidence on inflation expectation formation. *Journal of Economic Dynamics and Control*, *44* (C), 147–168.
Pfajfar, D., & Zakelj, B. (2018). Inflation expectations and monetary policy design: Evidence from the laboratory. *Macroeconomic Dynamics*, *22* (4), 1035–1075.
Potters, J., & Suetens, S. (2009). Cooperation in experimental games of strategic complements and subsitutes. *Review of Economic Studies*, *76* (3), 1125–1147.
Rholes, R., & Petersen, L. (2021). Should central banks communicate uncertainty in their projections?. *Journal of Economic Behavior and Organization*, *183* (C), 320–341.
Smith, V. L., Suchanek, G. L., & Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica*, *56* (5), 1119–1151.
Stöckl, T., Huber, J., & Kirchler, M. (2010). Bubble measures in Experimental Asset Markets. *Experimental Economics*, *13* (3), 284–298.
Sutan, A., & Willinger, M. (2009). Guessing with negative feedback: An experiment. *Journal of Economic Dynamics and Control*, *33* (5), 1123–1133.
Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.

Cite this article: Hanaki, N., & Takahashi, Y. (2025). An experiment on a multi-period beauty contest game. *Experimental Economics*, 28(6), 1214–1239. https://doi.org/10.1017/eec.2026.10048

https://doi.org/10.1017/eec.2026.10048 Published online by Cambridge University Press