



Review article

Expectation formation in finance and macroeconomics: A review of new experimental evidence[☆]Te Bao^a, Cars Hommes^{b,c}, Jiaoying Pei^{d,*}^a School of Social Sciences and NTU-WeBank JRC, Nanyang Technological University Singapore, Singapore^b CeNDEF, School of Economics, University of Amsterdam, Netherlands^c Bank of Canada, Canada^d School of Social Sciences, Nanyang Environment & Water Research Institute, Environmental Process Modelling Centre, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:

Received 24 November 2020

Received in revised form 17 September 2021

Accepted 21 September 2021

Available online 24 September 2021

JEL classification:

C10

D17

D84

E52

G12

G17

G40

Keywords:

Learning to forecast experiment

Experimental finance

Rational expectations

Bubbles and crashes

Behavioral finance

ABSTRACT

This paper reviews the recent development and new findings of the literature on learning-to-forecast experiments (LtFEs). In general, the stylized finding in the typical LtFEs, namely the rapid convergence to the rational expectations equilibrium in negative feedback markets and persistent bubbles and crashes in positive feedback markets, is a robust result against several deviations from the baseline design (e.g., number of subjects in each market, price prediction versus quantity decision, short term versus long term predictions, predicting price or returns). Recent studies also find a high level of consistency between findings from forecasting data from the laboratory and the field, and forecasting accuracy crucially depends on the complexity of the task.

© 2021 Elsevier B.V. All rights reserved.

Contents

1. Introduction.....	2
2. Basic setup of a learning to forecast experiment.....	3
2.1. Experimental design.....	3
2.2. Price dynamics and individual expectations.....	3
3. Stylized results from recent literature.....	4
3.1. LtFE versus LtOE.....	4
3.2. Large scale LtFEs.....	5
3.3. Time horizon.....	6
3.4. Price versus return.....	6
3.5. Monetary policy experiments.....	7
3.6. Laboratory experiments and computational experiments.....	8
3.7. Comparing expectation data from the lab and the field.....	8

[☆] We thank the editor, an associate editor and an anonymous referee for their helpful and detailed comments. We benefitted from stimulating discussion by Nobuyuki Hanaki and Luba Petersen. Te Bao thanks the financial support from the AcRF Tier 1 Grant RG69/19 from Ministry of Education of Singapore, NTU-WeBank Joint Research Center (NWJ-2019-001, NWJ-2020-003) and the National Science Foundation of China (No. 71803201, 71773013, and 71873149).

* Corresponding author.

E-mail addresses: baote@ntu.edu.sg (T. Bao), c.h.hommes@uva.nl (C. Hommes), peij0003@e.ntu.edu.sg (J. Pei).

3.8. Cognitive ability, task complexity and experience 9
 4. Conclusion 10
 References 11

1. Introduction

Expectation formation plays a significant role in modern finance and macroeconomic modeling. Since the circulation of seminal works by Muth (1961) and Lucas (1972), the rational expectations hypothesis (REH) has been the standard approach to model expectation formation. However, due to the lack of high-quality observational data on agents' expectation formation and the difficulty for "testing joint hypotheses", it is usually difficult to set up a clean test on the REH using empirical data from the field.

In recent years, learning-to-forecast experiments (LtFEs), an experimental design that dates back to Marimon and Sunder (1993, 1994, 1995) and Marimon et al. (1993) has been widely used by experimental economists to study expectation formation in financial markets and macroeconomies. The key feature of a learning-to-forecast experiment (LtFE) is that the subjects of the experiment play the role of professional forecasters (Hommes, 2011, 2013b, 2021). Their only task is to submit their expectation on an economic variable, e.g., the market price, the inflation rate, or the output gap. After collecting individual expectations, the conditional optimal quantity decision (e.g., trading, investing, and saving) are calculated by a computer algorithm, which aids in determining the realization of the variables on which the subjects made their forecasts. The learning-to-forecast (LtF) approach is usually contrasted with the alternative learning-to-optimize experiment design (LtOEs Duffy, 2010, 2014; Arifovic and Duffy, 2018), where the subjects simply make their choice decisions. Depending on the context, the choice decision may refer to various quantity or trading decisions, e.g., the consumption or saving decision in an intertemporal choice problem for a household in, e.g., Lei and Noussair (2002); the supply quantity decision for a firm in, e.g., Bao et al. (2013); and a bid or ask made by a trader in a double auction market in, e.g., Smith et al. (1988) and surveyed by Palan (2013). Because the LtFE design directly elicits and incentivizes the individual expectations, the subjects should have stronger incentive to form rational expectations (RE).

The market can display positive or negative feedback. The asset markets are considered positive feedback systems, where the realized market price increases when individual price forecasts increase. A classical cobweb framework describing a supply-driven commodity market with a production lag, on the other hand, exhibits negative feedback—that is, a higher expected price leads to increased production and thus to a lower realized market price (Hommes, 2013a, 2021). A general conclusion from the LtFE literature is that the agents can learn the rational expectations equilibrium (REE) when the market is a negative feedback system (e.g., Hommes et al., 2000). Yet, agents fail to learn the REE when the market is a positive feedback system (e.g., Hommes et al., 2005, 2008). There have been several comprehensive surveys of this existing literature (e.g., Hommes, 2021, 2011, 2013a,b, 2014; Assenza et al., 2014). In this paper, we focus on the relatively new development in this literature, i.e., studies published in the 2010s, to summarize recent trends and discuss possible future directions for research in this field. The new designs and research questions of papers surveyed in this paper mainly fall into the following categories:

- (1) Experiments that compare the learning-to-forecast (LtF) and the learning-to-optimize (LtO) design (e.g., Bao et al., 2013, 2017; Mirdamadi and Petersen, 2018; Giamattei et al.,

- 2020). The main result of this strand of literature is that all things equal, the convergence to the REE is less likely or slower when in LtOEs than in LtFEs.
- (2) Like other market experiments, the market size of a typical LtFE is usually 6–10 participants. In recent years, researchers start to run large-scale LtFEs (e.g., Bao et al., 2020) to test if the results from relatively small-scale experiments are robust in larger experimental markets. These studies usually find that bubbles and crashes are still prevalent in markets consisting of more subjects.
- (3) A typical LtFE usually lasts for 50 periods, and the predictions are made one period or two periods ahead. In recent years, researchers have started to run LtFE with longer horizons to investigate the role of long-run predictions (Colasante et al., 2018; Evans et al., 2019; Anufriev et al., 2020a,b). The findings suggest that markets tend to be more stable when people make long-run instead of short-run expectations and when the length of the experiment is greater (i.e., with more periods).
- (4) Traditionally, LtFEs on asset market elicits beliefs on asset prices. Some recent studies compare cases where agents form expectations on prices versus returns (Glaser et al., 2019; Hanaki et al., 2019b). The main result of this strand of literature is that all things equal, subjects' expectation is higher and bubbles are more likely to form when they predict in terms of returns instead of prices.
- (5) Traditionally, LtFEs mainly study questions related to asset pricing. Recent LtFEs pay more attention to monetary economics and the role of monetary policy in asset markets (e.g., Arifovic and Petersen, 2017; Arifovic et al., 2019; Assenza et al., 2019; Bao and Zong, 2019; Hommes et al., 2019a,b; Mauersberger, 2019; Ahrens et al., 2019). The findings of this literature show that higher interest rates and central bank communication are useful in stabilizing asset prices or inflation rates.
- (6) Certain studies have tried to compare data on expectation formation from the lab and from the field (e.g., Landier et al., 2019; Cornand and Hubert, 2020). The results of this literature show a high consistency between the findings of lab experiments and data from the field.
- (7) Other papers have attempted to combine laboratory and computational experiments (e.g., Hommes et al., 2017; Anufriev and Hommes, 2012; Bao et al., 2012; Anufriev et al., 2016, 2018, 2019a). The results of this literature usually show that subjects' expectation formation is usually better explained by computational economics models where subjects choose from a menu of simple heuristics.
- (8) Finally, there exist studies on how the complexity of the decision and subjects' cognitive ability and experience influence forecasting behavior (e.g., Mirdamadi and Petersen, 2018; Anufriev et al., 2019a; Arifovic et al., 2019; Bao and Duffy, 2021; He and Kucinskas, 2019). The results of this literature demonstrate that people tend to use simple heuristics even when faced with highly complex tasks where multiple equilibria are possible. Higher cognitive ability of the subjects leads to better convergence to the REE, but debate persists about whether experience helps with convergence and market stability.

The rest of the paper is organized as follows: in Section 2, we go through the basic setup of a LtFE. Next, in Section 3, we list the

main results of the recent literature. Specifically, we discuss the literature on each of the eight types of studies in each subsection of Section 3—that is, in Section 3.1 we discuss the literature on type 1, while Section 3.2 covers the literature on type 2. In Section 4, we draw a short conclusion and offer discussion based on the development of the literature.

2. Basic setup of a learning to forecast experiment

2.1. Experimental design

A baseline LtFE is usually a market experiment with 6–10 subjects in each market. This type of experiment usually employs a between-subject design. One market serves as one independent observation. The subjects make their forecast on one or two economic variables, e.g., the price of a product/financial asset, the inflation rate, GDP output gap, etc. To provide the appropriate incentive to do their best in making an accurate forecast, their payoff is a decreasing function of their prediction error. Some studies make the subjects' payoff a quadratic loss function of their prediction error, while others put prediction error in the denominator of the payoff function.

In a LtFE, the subjects usually play the role of a professional consultant/forecaster/analyst of a firm. Their expectations are fed into the decision problem of the firm in determining their output/trading/investment decisions. Other things equal, a more accurate prediction is associated with a higher profit for the firm and better compensation to the subject.

A LtFE is usually a multi-period experiment. The subjects need to predict the economic variable for 40–65 consecutive periods. In each period, their information set usually includes the history of their own past predictions and the realization of the economic variable. They usually do not know the data generating process (DGP) of the economic variable, as most market participants do not know the DGP of GDP, stock prices, or the inflation rate in real life. The experiment usually uses the simultaneous decision setting, which means that they do not have information on others' expectations in the same period, and still do not have it even after the realization of the economic variable is revealed. In a way, a LtFE differentiates from market experiments with strategic substitutes and complements (Fehr and Tyran, 2005, 2008) in that it is not a game between the subject and other players as his/her opponents, but a game between the subject and “the market”. Thus, a subject in a LtFE is usually considered a price taker, who does not put much consideration on his/her market power in the decision-making process.

A key research question of the LtFE literature is this: when people do not start from the REE and do not have the knowledge about the specification of the DGP of the economy, can they learn to play RE over time? Stated differently, can learning lead the market to converge to its REE? According to the REH, this should be the case. Instead of assuming that every agent has full information about the economy, the theoretical prediction by REH is that people should be able to learn the REE as long as they have the incentive to search for information and try to form an accurate forecast. Meanwhile, their prediction errors should not have a cross-sectional correlation.

2.2. Price dynamics and individual expectations

Fig. 1 shows the aggregate price dynamics in a typical LtFE. While the markets with negative feedback usually converge to the REE (dashed line) within five periods after the experiment starts, or after the experimental economy experiences a large exogenous shock, markets with positive feedbacks usually fail to converge to

the REE and exhibit prolonged oscillations and deviation from the underlying REE/fundamentals.

To better understand the individual expectation formation in the experimental markets of LtFEs, researchers use different methodologies to categorize the forecasting behavior by individual subjects in these markets. One important behavioral model used in this literature is the heuristic switching model (HSM) by Anufriev and Hommes (2012).

The basic setup of an HSM is that in each period, the subjects choose from a menu of forecasting heuristics. They can observe the history of the forecasting accuracy of each heuristic, and the key assumption of the model is that the heuristics that perform better in terms of generating smaller forecasting errors in the recent past are assigned with higher evolutionary fitness; hence, they attract more followers in the next period. There are typically four forecasting strategies in an HSM:

An adaptive expectations rule (ADA), $p_{h,t}^e = p_{t-1}^e + w(p_{t-1} - p_{t-1}^e)$, where the prediction is a weighted average of the previous prediction and the last observed price.

A weak trend following (WTR) rule (or a contrarian rule, CTR), $p_{h,t}^e = p_{t-1} + \gamma(p_{t-1} - p_{t-2})$, $\gamma \in (0, 1]$ (or $\gamma < 0$ for a contrarian rule), where the prediction is the last observed price plus the last observed price change multiplied by a constant parameter between 0 and 1 (less than 0 for a contrarian rule).

A strong trend extrapolation (TRE) rule, $p_{h,t}^e = p_{t-1} + \gamma(p_{t-1} - p_{t-2})$, $\gamma > 1$, where the prediction is also the last price plus the last observed price change multiplied by a constant parameter greater than 1.

An Anchoring and Adjustment rule (A&A, Tversky and Kahneman, 1974), $p_{h,t}^e = 0.5(p_{t-1}^{av} + p_{t-1}) + (p_{t-1} - p_{t-2})$, which uses a time varying anchor, or the average of the last price and the sample mean of all past prices, $(p_{t-1}^{av} + p_{t-1})$, then extrapolates the last price change, $(p_{t-1} - p_{t-2})$.

Let $n_{h,t}$ be the fraction of subjects using heuristic h in period t . The specific weight updating rule is given by a discrete choice model with asynchronous updating:

$$n_{h,t} = \delta n_{h,t-1} + (1 - \delta) \frac{\exp(\beta U_{h,t-1})}{\sum_{i=1}^4 \exp(\beta U_{i,t-1})}$$

δ is the parameter that captures the inertia people stay with the previous heuristic. The parameter β represents the “sensitivity” to switch. The higher the β , the faster the participants switch to more successful rules in the recent past. $U_{h,t-1}$ is a fitness measure that is decreasing in the forecasting heuristic's forecasting error.

As shown in Fig. 2, individuals usually follow adaptive (ADA) or contrarian (CTR) expectations in negative feedback markets, and a strong trend-following (TRE) rule or an anchoring and adjustment (A&A) rule in positive feedback markets. They are thus able to converge to the REE using adaptive expectations, especially when the markets are E-stable (Evans and Honkapohja, 1999, 2003, 2009) in negative feedback markets. Still, they are usually unable to learn the RE equilibrium in markets with positive feedbacks, because riding on a common trend leads to violation of “uncorrelated prediction errors” across individuals.

More recently, Bao and Hommes (2019) studied the price dynamics in experimental housing markets as a “hybrid” of positive and negative feedback systems. The housing market is a production market (as a negative feedback system for the builders), and an asset market (as a positive feedback market for the speculators). The result of the experiment shows that the market price tends to be more stable when the “strength” of negative feedback, i.e., the slope of the supply function is larger. The result provides supportive evidence that all things equal, housing markets with larger supply elasticity should experience fewer bubbles and crashes. These results also show that overall weak positive feedback leads to a stable market, while strong positive feedback creates bubbles and crashes.

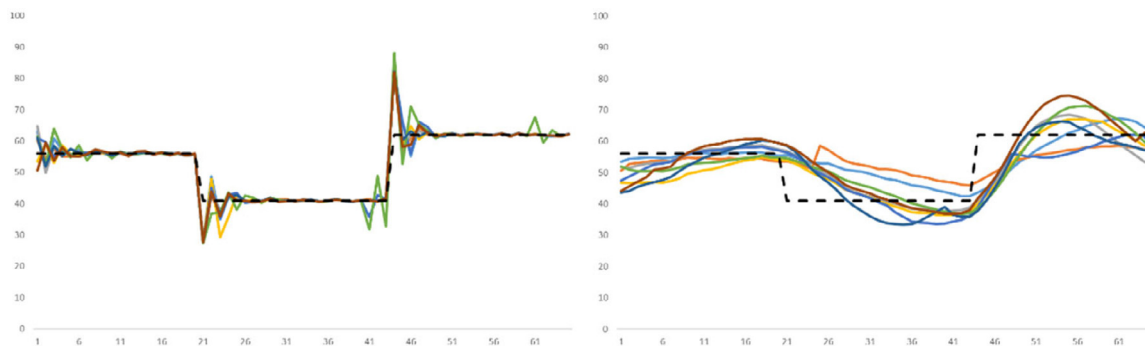


Fig. 1. Price dynamics in negative (left panel) and positive feedback (right panel) markets in the LtFE by Bao et al. (2012).

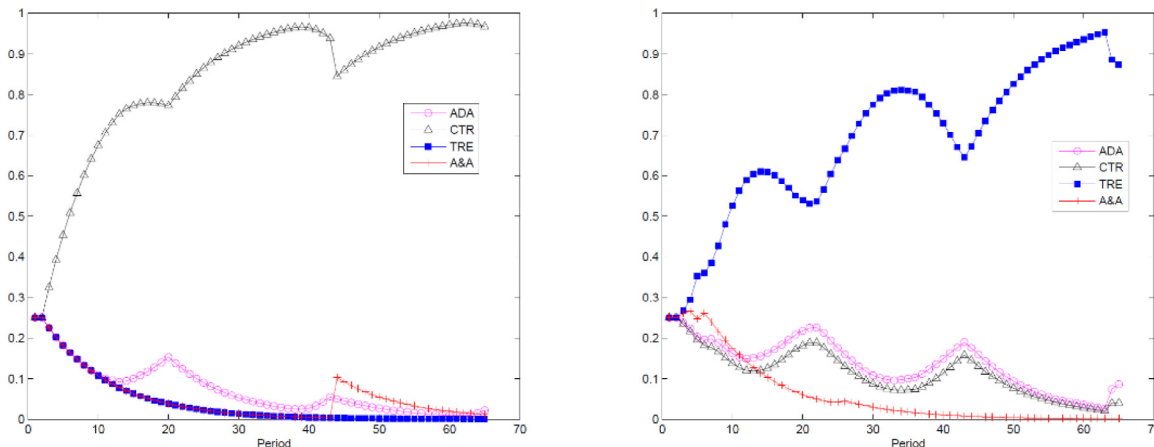


Fig. 2. Simulated fractions of individuals using different forecasting strategies in a typical negative feedback market (left panel) and a positive feedback market (right panel) in the LtFE by Bao et al. (2012).

3. Stylized results from recent literature

In this section, we review the results of some recent studies (most of which were conducted or published after 2010) in the LtFE literature. We do not attempt to cover all details of the design and results of all studies, but instead highlight the main conclusions and supporting evidence.

3.1. LtFE versus LtOE

Observation 1: The convergence to the REE is not more likely or faster when the subjects submit quantity decisions instead of making price forecasts. Rather, convergence may be slower, and bubbles and crashes are still prevalent under quantity decisions.

Support: Since the emergence of the earliest LtFE literature, there have been questions about the comparability between the results from LtFEs and LtOEs, where subjects make quantity decisions directly. Though there have been some LtOEs that also elicit price forecasts (e.g., Cheung et al., 2014; Cohn et al., 2015; Haruvy et al., 2007; Hanaki et al., 2018), the price forecast in those experiments is more like a by-product of the experiment: it does not enter the DGP of the market price, and hence plays a minimal role in the experiment as opposed to expectation formation in LtFEs.

To our knowledge, Bao et al. (2013) is the first experiment that sets up comparable LtFE, LtOE treatments, as well as the combination of the two. There is a shared cobweb economy model in all treatments, where the subjects play the role of advisors of competing companies producing consumer products. The good is an ordinary good, so that demand is a downward-sloping function

of the price. The authors impose a quadratic cost function of production.

In the LtFE treatment, the subjects submit their price forecast in each period. The price is then determined by the average price forecast, and the subjects are paid according to their forecasting accuracy. In the LtOE treatment, the subjects submit their production quantity directly. The market price is then determined by the total supply quantity, and subjects are paid according to the profitability of this quantity decision. In a third treatment, they combine the two, the subjects submit both a price forecast and a production quantity. The market price is then determined by the total supply quantity as in the LtOE treatment, and subjects receive their payoff half from the forecasting task and half from the quantity decision task.

The result of Bao et al. (2013) shows that convergence is the fastest in LtFE and slowest in the combination of LtFE and LtOE. The authors further found that most subjects use adaptive rules to forecast prices. Given their price forecast, subjects fail to provide the conditionally optimal quantity in the treatment with both forecasting and optimizing tasks. The results suggest that LtFE indeed provides an “upper bound” of how well the REH works in markets.

Unlike Bao et al. (2013), Bao et al. (2017) studied the expectation formation and price dynamics in positive feedback markets where subjects play the role of advisors for investment companies. In Bao et al. (2017), the company will buy more assets if the subject’s prediction of the future asset price is higher. The authors also designed three treatments: LtFE, LtOE, and a third one called Mixed, where the subjects perform both forecasting and quantity decision (on trading) tasks. To avoid potential hedging, the subjects in the Mixed treatment receive their payment based on

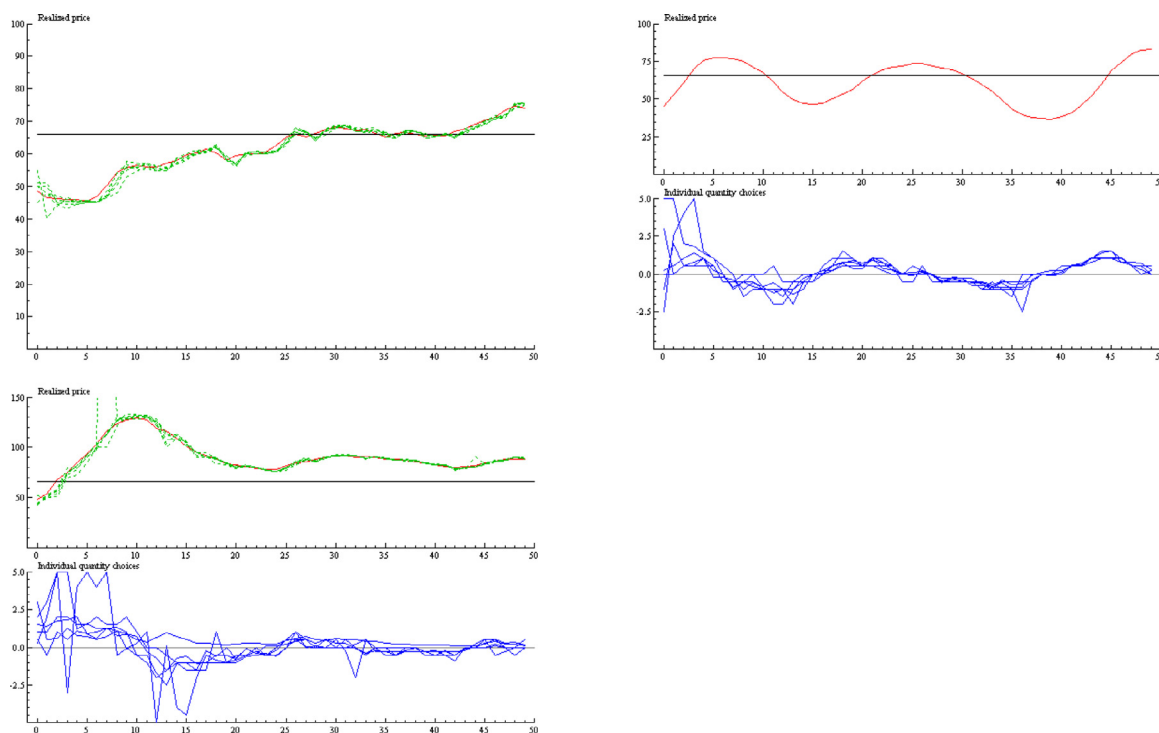


Fig. 3. The asset price in a typical market in the LtFE (top left panel), LtOE (top right panel), and Mixed (bottom panel) in Bao et al. (2017).

their performance in the forecasting and trading task with 50:50 probability, instead of 50:50 weight.

Fig. 3 presents the asset price dynamics in a typical market in the LtFE, LtOE, and Mixed in Bao et al. (2017). None of the markets converge to the REE, but between treatments, the price deviation and the magnitude of fluctuation are significantly larger in the LtOE and Mixed treatments than in the LtFE treatment.

The authors also tried to provide an empirical micro-foundation of observed differences across the three treatments. They estimated individual forecasting and trading rules and found significant differences across treatments. In the LtFE treatment, individual forecasting behavior is more cautious. Subjects use a more conservative anchor (a weighted average of last observed price and last forecast) in their trend-following rules. In contrast, in the Mixed treatment almost all weight is given to the last observed price, leading to a more aggressive trend-following forecasting rule. Individual trading behavior of most subjects can be characterized by extrapolation of past and/or expected returns, and the degree of the return extrapolation coefficients are higher in the LtOE and Mixed treatments.

Arifovic et al. (2019) built an overlapping generations (OLG) economy of Grandmont (1985). Their LtFE treatment offered the first example of spontaneous coordination on a two-cycle in the lab (as higher order cycle), while the LtOE treatment failed to converge to a two-cycle, even after the up and down oscillations induced by an initial training phase. The authors plotted the cumulative distribution of individual decision times and the length of instructions and reported a significantly higher cognitive load in LtOE than LtFE. In sum, they suggest the possibility that it is the strategic uncertainty or difference in cognitive load between the two designs that leads to the observed differences in outcomes.

Giamattei et al. (2020) have found that if subjects are asked to provide a price forecast on a double auction market, a la Smith et al. (1988), paying for the accuracy of the forecast tends to enlarge the mispricing and market instability. The reason may be that the incentive distracts the subjects' attention in tracking the fundamental value while trading.

3.2. Large scale LtFEs

Observation 2: Bubbles and crashes also occur in large experimental LtFE asset markets.

Support: Most standard LtFEs use the market size of 6 participants. Some may wonder if the results of this design are robust when the group size becomes larger. In particular, supporters of the REH may claim that RE works the best with a large economy populated by millions of people, and a large sample size may be a necessary and sufficient condition for “wisdom of crowds” to work. In this case, the large booms and busts in LtFEs with positive feedback may be mitigated or eliminated when the number of subjects in each experimental market grows larger.

In response to this question, a few recent LtFEs have used large-scale design, i.e., by increasing the market size from 6 to 20–30, or even to 100. These studies usually show that bubbles and crashes still occur in these large markets, just as they did in smaller markets.

Bao et al. (2020) studied the price dynamics and individual expectations in LtFE markets. In Fig. 4, each solid line represents one market in the experiment. The experimental setup is the same as in Hommes et al. (2008), except that the market size increases from 6 to 21–32. The unique REE of the market price is 60 (dashed line), but the results show that similar to markets in Hommes et al. (2008), 6 out of 7 markets show persistent divergence from the REE, and the peak of the price cycle can be as high as almost 1000. Thus, the findings in Hommes et al. (2008) are robust when the market size increases from 6 to 20–30.

The price dynamics in the seven markets from Bao et al. (2020) are shown in Fig. 4. As the figure shows, the price dynamics follow the same pattern as in Hommes et al. (2008), and there is no evidence that a larger group size reduced the size or likelihood of bubbles.

Hommes et al. (2021) further extended the size of the large experimental asset market to around 100 subjects (between 92 and 104) in each market. The unique REE of the asset price in this

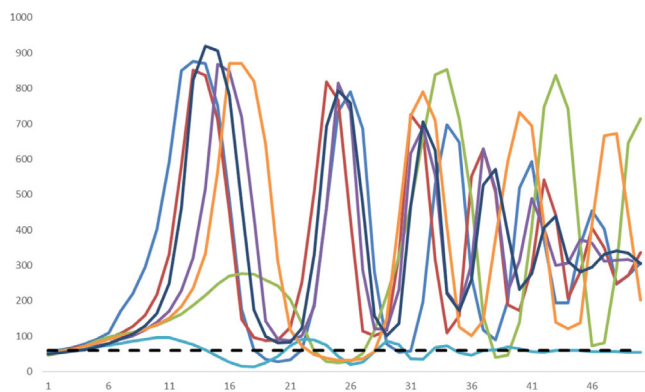


Fig. 4. Price Dynamics in Bao et al. (2020). The solid lines are market prices in each experimental market, and the dashed line is the REE.

experiment is 66. The average asset price is 139.38 for the Large groups and 153.41 for the Small groups (with six subjects each as in a standard LtFE). While the overvaluation seems smaller in large markets, it is still far from zero, and large bubbles occur in 3 out of 6 large markets. The authors also examined the effect of news announcements when the market is overheated and find that it can help to substantially bring down asset prices.

3.3. Time horizon

Observation 3: Markets populated with more long-run forecasters are more likely to converge to the REE. Long-run forecasters' forecast is better described by adaptive learning, while short-run forecasters are usually trend-extrapolators.

Support: Evans et al. (2019) ran a LtFE where subjects play the role of agents with CRRA utility functions and solve a consumption-based asset pricing problem a la Lucas Jr. (1978). In this setting, a boundedly rational agent model by Branch et al. (2012) proposes that when agents make “T-period ahead optimal learning”, the asset price will converge to its REE faster when T is larger. The authors designed four treatments where the market is populated by 0%, 30%, 50%, and 100% of long-run forecasters who make ten periods ahead forecasts (while the rest are subjects who make one period ahead forecasts as in standard LtFEs). The result shows that short-horizon markets are prone to persistent deviations from RE. By contrast, markets populated by even a modest fraction of long-horizon forecasters exhibit convergence towards the REE. Long-horizon forecasts are well-described by adaptive learning, which leads to convergence and stabilization, while short-horizon forecasts are typically users of destabilizing trends following strategies.

Parallel to the paper mentioned above, Anufriev et al. (2020a) examined how long-run expectations influence market stability in markets with positive expectation feedback. Different from Evans et al. (2019), their experimental setting is the standard after Brock and Hommes (1998) and Hommes et al. (2005, 2008). In this study, long-run expectation means the subjects can make two periods ahead or three periods ahead expectations, and there is no treatment with a mixture of short-run and long-run forecasters, i.e., all subjects face the same forecasting time horizon in each market. The authors also introduced the initial history of past prices at the beginning of the experiment. That is, instead of seeing no past prices, the subjects can observe a long history of asset prices from markets in previous asset pricing LtFEs. Like Evans et al. (2019), the result of this paper shows that long-run expectations tend to help stabilization and convergence to REE. All markets that start with the history of converging prices tend

to remain stable. For the markets that start with the history of oscillating price dynamics, the price tends to be more stable when the subjects make the long-run instead of short-run expectations.

There are other studies that elicit long-run expectations together with short-run expectations, e.g., Colasante et al. (2018, 2020). But since the long-run expectations in those experiments do not enter the DGP of the realized asset prices, they play a lesser role in the experiment and tend to generate a smaller impact.

Observation 4: Increasing the length, i.e., the number of periods, and time pressure can help the markets to converge to the REE.

Support: A typical LtFE is a 50-period experiment. Some wonder how the price dynamics will look if the number of periods increases, i.e., whether those markets that do not converge in the first 50 periods will converge after 50 periods. To address this issue, Anufriev et al. (2020b) ran a LtFE with positive expectation feedback where the length increases by a factor three, i.e., to 150 periods. The study showed that the result may go both ways. Some markets do not converge in the first 50 periods but converge afterward, while there are also markets that seem to be stable in the first 50 periods but start to oscillate around the end of the experiment. Overall, more markets fall into the first category. Increasing the length of the experiment does seem to help to stabilize the market.

In the same experiment, the authors also varied the time pressure faced by the subjects. The subjects had 25 s to make their decision in the low time pressure treatment, but only 6 s in the high time pressure treatment. The authors found that the subjects are somehow less trend-chasing under high time pressure, which helps to stabilize the market.

3.4. Price versus return

Observation 5: The format in which the data is presented, or in which the prediction is submitted, does impact forecasting behavior. All things equal, subjects' expectation is higher, and bubbles are more likely when they predict in terms of returns instead of prices. The results on the effect of the format of past data are mixed, while some studies find that price expectation tends to be lower, and bubbles are less likely when the past data is presented in terms of returns instead of prices. Other studies found no significant effect.

Support: Both expectations and past data on financial variables can be returns or price levels in the real world. Glaser et al. (2019) studied whether the format of expectation has an impact on expectation formation; they found that on average, the expectations are higher by between 1.1 and 2.4 percentage points per month if subjects predict returns rather than price levels. In contrast, showing subjects with return bar charts as opposed to price line charts leads to a lowered expectation by 1.7 to 1.0 percentage points per month. This finding is robust against whether the payoff for the subjects is fixed or performance-based across different subject pools (i.e., students or financial professionals).

Glaser et al. (2019) is an experiment on expectation formation using an exogenously generated price time series. While it can provide a good description of expectation formation behavior at the individual level, it is difficult to conclude how the patterns of behavior influence aggregate market stability. Hanaki et al. (2019a) conducted a LtFE where the subjects' price or return forecast will be a key variable in determining the asset prices and returns. They used a two-by-two design where the two dimensions are (1) if subjects predict the prices or returns, and (2) if subjects observe information about the past in terms of prices or returns. The paper shows that while the price bubble is again larger when subjects predict returns compared to when they predict prices, there is no evidence that the format of how past information is presented influences price dynamics and market stability.

3.5. Monetary policy experiments

Observation 6: Though LtOEs usually find little or no evidence of the effectiveness of higher interest rates to curb asset bubbles, LtFEs on asset markets usually find supportive evidence for the effectiveness of monetary policies. Central bank communication helps stabilize expectations if it is done in a simple and accessible way.

Support: In the LtOE literature, [Fischbacher et al. \(2013\)](#) was the first experimental study on the impact of monetary policy on asset prices in double auction markets. They found that a higher interest rate leads to lower liquidity in the market but has little impact on the level of the asset prices in double auction markets, as [Smith et al. \(1988\)](#). Similarly, [Giusti et al. \(2016\)](#) concluded that introducing the opportunity cost of speculation in the form of interest payment to cash has limited success in mitigating bubbles.

[Bao and Zong \(2019\)](#) conducted a LtFE where the REE of the asset price is 60. The initial interest rate for the risk-free asset is 5%. They designed three treatments:

Treatment B: the baseline treatment where the interest rate is unchanged over time.

Treatment P: The central bank will raise the interest rate to 10% if the asset price in the previous period is higher than 90 (50% higher than the REE); they will lower it to 2.5% if the asset price in the previous period is lower than 30 (50% lower than the REE). The subjects are informed about this policy in the instructions.

Treatment PN: This treatment is the same as treatment P, except while subjects see the real-time interest rate in each period, they are not informed about the detailed scheme and purpose of the policy.

[Fig. 5](#) shows the asset price dynamics in each of the eight markets in Treatment B (left panel) and Treatment P (right panel) in [Bao and Zong \(2019\)](#). The result of the experiment shows that the introduction of the monetary policy can reduce by two-thirds the relative absolute deviation (RAD, a commonly used measure of price bubbles in the experimental finance literature proposed by [Stöckl et al. \(2010\)](#)) of the market price from the fundamental value of the market. Moreover, the effectiveness of the policy does not depend on whether the subjects are informed about the purpose of the policy. The result of [Bao and Zong \(2019\)](#) suggests that a higher interest rate does help to curb the “bubbly” price expectations and therefore reduces mispricing. The reason why such policies did not work in LtOEs is more likely related to bounded rationality in subjects’ quantity decision making in trading, not to the expectation formation process.

Parallel to [Bao and Zong \(2019\)](#), [Hennequin and Hommes \(2019\)](#) studied how a Taylor-rule-like interest policy can reduce asset bubbles. The interest rate policy is a linearly increasing function of the price deviation from the fundamental value in their policy treatments. Depending on the strength of the policy, they further differentiated between a weak and a strong rule treatment. In their weak rule treatment, the interest rate will increase by 0.001% when the price deviation from the fundamental value increases by 1%. In their strong rule treatment, the interest rate will increase by 0.1% when the price deviation from the fundamental value increases by 1%. The result shows that while the weak rule does not stabilize the market, the strong rule can reduce the price deviation by 67%–90%, similar to the effect found in [Bao and Zong \(2019\)](#).

[Assenza et al. \(2019\)](#) tested the effectiveness of the Taylor Principle using a self-referential LtFE in the New Keynesian framework. Their result suggests that when demanding a convergence towards the inflation target, the Taylor Principle is a necessary, but not a sufficient condition. Instead, central banks need to use an aggressive enough monetary policy rule by introducing strong enough negative feedback between expected

inflation and aggregate demand. A sufficiently strong Taylor rule can manage expectations because the policy avoids coordination on trend-following behavior and prevents expectation errors from becoming self-fulfilling. Similarly, [Kryvtsov and Petersen \(2013\)](#) also found that the Taylor rule monetary policy proved a highly effective device in lowering the conditional variance of output gap and inflation. [Mauersberger \(2019\)](#) ran a LtFE in a New Keynesian Economy as in [Woodford \(2013\)](#) and found that the Taylor principle does not necessarily guarantee convergence to the steady state, but the welfare loss due to expectation-driven volatility can be largely mitigated by the Taylor principle near the steady state.

[Kryvtsov and Petersen \(2021\)](#) studied if central bank communication can stabilize individual forecasts and aggregate outcomes. They conducted a LtFE based on an extended version of [Woodford \(2013\)](#) model of heterogeneous expectations and monetary policy. The output in the economy is subject to an AR(1) demand shock. Subjects are aware that the central bank responds to deviations of inflation and output gap from target, and that the central bank reacts more than one-for-one with inflation. There are four treatments in their experiment: (1) The control treatment with no communication; (2) COM-BACK treatment, where the central bank simply makes an announcement on whether the interest rate has increased, decreased, or stayed unchanged in the last period. Note that by default, participants in all treatments can observe the history of interest rates in the experimental interface. This treatment does not provide new information, but just increases the salience of the information; (3) COM-FWD treatment, in which all subjects receive an announcement on the central bank’s expected policy decision to increase or decrease the interest rate, or to let it stay unchanged. Subjects are informed about the function used by the central bank to forecast future interest rates; (4) COM-COMMIT treatment, where the central bank will occasionally let the nominal interest rate remain unchanged for some periods and inform the participants in advance about whether the interest rate will be changed in the next periods. The authors found that the fluctuation of the economy is smaller in all treatments with communication than in the control treatment. The reduction in individual expectations and the aggregate outcome is the largest in COM-BACK treatment, suggesting that communication is more effective if done in a simple and relatable backward-looking way.

Observation 7: It is difficult to escape the liquidity trap using monetary policy alone. Monetary policy can lead the economy to the targeted steady-state equilibrium when combined with fiscal policy. Publishing strategic central bank projections may help the economy to escape the liquidity trap if the central bank can gain sufficient credibility from the private sector investors.

Support: Two works use LtFE to study how the economy can escape from a liquidity trap when the interest rate is near the zero-lower-bound ([Arifovic and Petersen, 2017](#) and [Hommes et al., 2019a](#)). Both experiments are based on a New Keynesian economy where individuals form expectations on future inflation rates and output gap and are paid according to their forecasting accuracy. In this economy, there are two equilibria—the target equilibrium and a low inflation equilibrium under RE, referred to as the zero lower bound (ZLB) steady state. [Evans et al. \(2008\)](#) showed that the target equilibrium is stable, while the low inflation equilibrium is an unstable saddle point under adaptive learning.

There are several differences between the two experiments: first, [Hommes et al. \(2019a\)](#) used the nonlinear NK model, while [Arifovic and Petersen \(2017\)](#) used the linear approximation of the model. Arifovic and Petersen introduced autocorrelated shocks to the system, while [Hommes et al. \(2019a\)](#) used expectational shocks generated by news announcements.

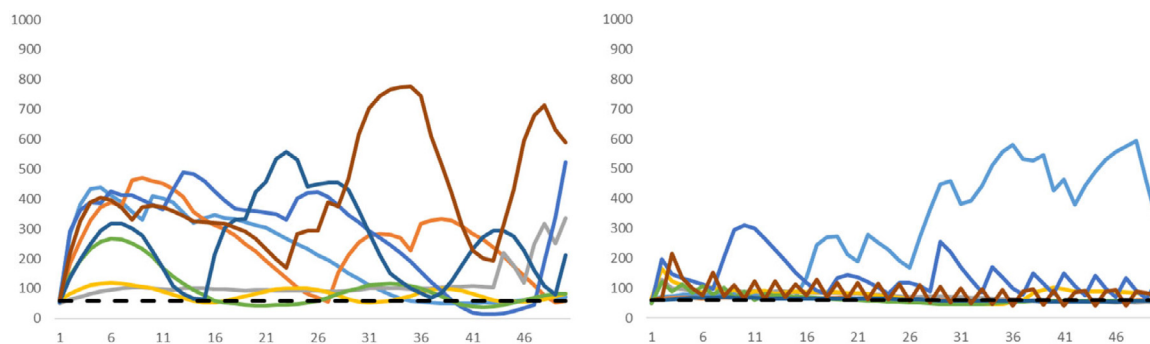


Fig. 5. The asset price in each of the eight markets in Treatment B (left panel) and Treatment P (right panel) in Bao and Zong (2019).

Despite the differences in the design, the two experiments reached similar conclusions. They both found that it is difficult to stabilize the economy and evade the deflation spiral using monetary policy alone. The monetary policy only works when combined with fiscal policies like “fiscal switching” (Chung et al., 2007). The results of both experiments are well in line with the adaptive learning model for expectations.

Ahrens et al. (2019) studied whether central banks can manage private-sector expectations by means of publishing one-period ahead inflation projections in a New Keynesian learning-to-forecast experiment. Their experimental economy is similar to Assenza et al. (2019), except that they introduced negative expectational shocks to the economy in three consecutive periods that may lead the economy to a deflationary spiral. In Treatment 2 and 3 of their experiment, the central bank of the economy, played by a human subject (Treatment 2) or a computer algorithm (Treatment 3), can publish inflation projection that serves as additional public information to the subjects. In Treatment 2, the human central banker received two forecasts from the computer: one was a data-driven forecast that would most likely prevail in the next period (i.e., with the smallest expected prediction error) and the other was a strategic, “required for target” forecast that could help the economy jump out of the deflationary spiral if all private sector agents follow the forecast. The central banker has the incentive to manage expectations so that the private sectors believe in and copy the strategic forecast, but to achieve this goal, the central bank must maintain good credibility by publishing projections that are not too far from realized inflation. In Treatment 3, a computer algorithm published the inflation projection automatically based on the tradeoff between expectation management and credibility. The result of the experiment shows that compared with no projection or with random projection, active central bank projections can drastically reduce the probability of deflationary spirals.

3.6. Laboratory experiments and computational experiments

Observation 8: It is difficult to explain subjects' expectation formation in lab experiments using the REH, or a single expectations formation rule. Subjects' expectation formation is usually better explained by computational economics models where subjects choose from a menu of simple heuristics, and these heuristics usually lead to a “smart” outcome for them, at least at the individual level. However, the aggregate market price may be subject to large and persistent bubbles and crashes due to temporary coordination on trend-extrapolating rules.

Support: The subjects' forecasting behavior is undoubtedly far from the REE in positive feedback markets. Though subjects may learn to play the REE in negative feedback markets, it is also

crucial to understand the learning path from the initial non-REE expectations to the REE. Recently, with the advancement of computational technologies and methodologies, researchers came up with computational models based on the evolutionary selection of forecasting heuristics to explain the experimental data. The two types of methods often used in the literature are the HSM based on Brock and Hommes (1998) model, and the Genetic Algorithm (GA) models.

The HSM (Anufriev and Hommes, 2012; Bao et al., 2012; Anufriev et al., 2016, 2018) is a relatively simple model to explain the expectation formation by subjects in LtFEs. The key assumption is that subjects choose from a small menu of forecasting heuristics (usually four), and the heuristic that performed better in the recent past will attract more followers in the future. This model has been successful in explaining expectation formation in both positive and negative feedback markets. The model is parsimonious because researchers only need to calibrate three parameters (the intensity of choice, memory, and inertia) of the model, and the results are indeed very robust for small changes in these parameters. The HSM suggests that people usually become more trend-chasing over time when playing in positive feedback markets and follow adaptive expectations in a negative feedback environment. Recently, Zhu et al. (2019) extended the HSM so that it can also be applied to LtOEs.

The more general GA model usually assumes that individuals search in a broad strategy space and switch between the strategies in a more sophisticated manner, trying to learn the parameters of the strategies over time. The literature goes back to Arifovic (1996, 1997) and Duffy (2006). Still, the large-scale application of the GA model to experimental data started in more recent years (Arifovic and Ledyard, 2011, 2012; Chen and Hsieh, 2011; Chen et al., 2011; Chen, 2012, 2013; Hommes and Lux, 2013; Chen et al., 2014; Hommes et al., 2017; Anufriev et al., 2013, 2019a,b; Tai et al., 2018; Makarewicz et al., 2020). With larger searching space and higher calculation capacity, the GA models can fit different moments of the experimental data and more detailed behavior at the individual subject level. The GA model can also provide an accurate estimate of the parameters in the first order heuristic widely used in LtFE and the coefficients for trend-chasing/contrarian behavior in HSMs.

3.7. Comparing expectation data from the lab and the field

Observation 9: Results based on experimental inflation forecasts data have a high level of external validity. Different sources of inflation forecasts (participants in experiments, households, industrial and financial professionals, and central bankers) share common patterns, and all deviate from the traditional RE paradigm.

Support: Cornand and Hubert (2020) carefully collected inflation expectations data from different sources, e.g., experimental data

from Pfajfar and Žakelj (2018), Cornand and M'baye (2018a,b), Adam (2007), Hommes et al. (2017) and survey data from Michigan household surveys, the Livingston survey on industrial professionals, Survey of Professional Forecasters and central bankers' forecast in FOMC meeting publication, and the Greenbook. They compared data from the lab and the field and find to find that they share a high level of common features. All deviate substantially from the RE hypothesis. The forecasting errors tend to be autocorrelated, and revision is made based on past information. These findings are in line with adaptive learning, as well as the "information rigidity" hypothesis proposed by Mankiw and Reis (2002, 2007) and Coibion and Gorodnichenko (2012, 2015).

Landier et al. (2019) and Bordalo et al. (2020) compared the expectation formation behavior by the subjects in a forecasting experiment where subjects predict an AR(1) time series with the field data from Coibion and Gorodnichenko (2015). They find that in both cases, the REH is firmly rejected. Subjects tend to overreact to recent trends and shocks, and the "forward-looking extrapolation" model can well explain the subjects' forecasting behavior.

Li (2020) used an online experiment to elicit subjects' expectations on future growth and inflation in China after the outbreak of COVID-19. He found that ambiguity-averse subjects tend to hold a more pessimistic view about the economic outlook in the future. Subjects seem to make consistent forecasting and consumption decisions. Those who predict lower growth also indicate that they are going to lower their consumption.

3.8. Cognitive ability, task complexity and experience

Observation 10: Subjects have bounded capability in handling complexity. Regardless of the complexity of the experimental environment, they consistently adopt simple belief-formation processes in LtFE.

Support: Arifovic et al. (2019) investigated the equilibrium selection under a complex OLG economy with multiple perfect-foresight equilibria, including periodic and chaotic dynamics. Theoretically, all equilibria can be selected under learning, provided that agents use a suitable rule. However, their experimental result shows that subjects would keep adopting the simple rules that are based on the information from the most recent observations, tracking low-order patterns instead of using higher order rules in the LtFE settings. The simple behavioral rules used by the subjects also lead to an aggregate convergence of prices and individual forecasts towards the simple equilibria even after a long transition, in this case the steady-state or the period-two cycle.

He and Kucinskis (2019) studied the effect of correlation on expectation formation, using a without-feedback LtFE framework to get rid of the ambiguous feedback effect on prediction accuracy. In their experiments, the subjects need to form expectations on an indicator A which, follows the AR(1) process. Subjects in the Baseline treatment observe only the past realized values of the indicator A. Subjects in the Correlated treatment additionally observe a leading indicator B that co-generates a bivariate VAR(1) with indicator A. In theory, the predictability of indicator A is the same for both treatments if subjects follow Bayesian updating when forming their expectations. The result of the study, however, shows that subjects predict with a significantly lower accuracy in the Correlated treatment. This is because when performing forecasting, subjects use a simplified mental model that largely ignores correlated variable.

Observation 11: There is mixed evidence on whether providing the subject full information on the structure of the economy can help to mitigate or eliminate the deviation from the REE.

Support: In most LtFEs, the subjects only have qualitative information on the underlying structure of the experimental economy.

One may wonder if it will be easier for the subjects to find the REE if they are provided full information, i.e., the exact equations of the price determination mechanism.

To our knowledge, the first experiment to address this issue is Sonnemans and Tuinstra (2010). Inspired by the convergence results in repeated beauty contest experiments and LtFEs with positive feedbacks (e.g., Hommes et al., 2005, 2008), they examined the key factors in determining whether a group of subjects can learn the REE. There are three main differences between a typical beauty contest game and a LtFE: (1) incentive structure (tournament incentives versus quadratic loss payoff function); (2) information structure (full information on the data generating process of the winning number versus limited information on the data generating process of the price); and (3) the feedback strength (2/3 versus 20/21). Their result shows that the difference in the experimental result is mainly driven by the feedback strength, while providing full information in LtFEs does not help much for subjects to learn to the REE price. Meanwhile, Bao and Duffy (2016) was one of the first LtFEs with negative feedback to introduce full information in their treatments. The research question of that paper is which of the adaptive learning model (Evans and Honkapohja, 2012) and the educative learning model (Guesnerie, 1992) provides a better explanation for people's price expectations in a simple cobweb market. Because educative learning requires that agents have complete knowledge on the data generating process (DGP) of the market price in the economy, Bao and Duffy (2016) showed the DGP to the subjects in the instructions in all treatment. Since there are no treatments without information of the DGP, it is impossible to draw conclusion about whether providing full information will facilitate convergence to REE in the experiment per se. On the other hand, if one compares the number of average periods before convergence in this experiment and other LtFEs with negative feedbacks, it seems the speed of convergence is similar in all LtFEs with negative feedbacks, i.e., the number of periods before convergence is usually fewer than 5. This result may serve as indirect evidence that given the high speed of convergence in LtFEs with negative feedbacks, it is difficult for treatments with more information to generate even faster convergence than the treatments with limited information in the traditional LtFE literature.

Mirdamadi and Petersen (2018) ran a LtFE where subjects form expectations on macroeconomic variables in a New Keynesian Economy. In their experiment, the eigenvalues of the economy are 0.88 and 0.67. So, the REE is a stable node under naïve expectations. They varied the level of the information the subjects receive on the data generation process of the economy. They found that providing precise quantitative training can help reduce the inflation forecast errors, reduce disagreements about inflation, and encourage a more substantial reaction to past forecast errors. Providing qualitative information shows a limited effect.

Multi-dimensionality is a common form of complexity faced by agents in macroeconomic or finance models. Anufriev et al. (2019a) extended the univariate LtFE into a planar system, using the setting of a beauty contest game (Nagel, 1995; Duffy and Nagel, 1997; Grosskopf and Nagel, 2008; Sutan and Willinger, 2009; Hanaki et al., 2019b) that provides subjects with full information about the DGP for the two endogenous variables. In particular, they focused on the simplest possible two-dimensional structure. As in a standard beauty contest game, variable a depends only on the average forecast for α . Variable b depends on the average predictions of both a and b . In their Saddle treatment with one unique converging path and an infinite number of other solution paths leading away from the steady state, subjects learn the steady state if it is a negative feedback market. Their result implies that with the full DGP of the planar system, participants

are able to learn and follow the saddle path in a system with negative feedback, even if they are not initially being placed on the saddle path itself. By contrast, subjects fail to coordinate when it is with positive feedback, just like in the univariate models.

The signal extraction model (DeGroot, 2004) is widely used in macroeconomics and game theory on how individuals form beliefs/expectations on the realization of economic variables based on two noisy signals. The main prediction of the theory is that the decision weight assigned to a signal is inversely related to its noisiness, as indicated by the variance of the distribution of the signal. The noisier the signal, the less decision weight. This theory was implicitly used in many game theory and finance experiments, e.g., the global game experiment on currency attack by Heinemann et al. (2004), but it was not tested directly. Bao and Duffy (2021) and Bao et al. (2019) tested the theoretical prediction in the laboratory. The subjects played the role of a financial advisor of an investment company. The company would be a buyer/seller of the asset if the subject's forecast is above/below the median forecast in the market. The authors found that on average, the subjects' prediction is explained well by the signal extraction model, though there is large heterogeneity in individual expectations. Subjects seem to apply some worst-case-scenario thinking and overestimate (i.e., take the upper limit of the distribution) the variance of an ambiguous signal whose variance is not a constant but varies between an interval.

Observation 12: Subjects with higher cognitive ability are more likely to form RE.

Support: Many studies (Akiyama et al., 2017; Zong et al., 2017; Bosch-Rosa et al., 2018) have shown that cognitive ability plays an important role in determining whether individuals can form RE. Akiyama et al. (2017) and Bosch-Rosa et al. (2018) ran LtOEs with call market design, and traders' expectation is elicited for the first period only while Zong et al. (2017) employed the LtFE design. In all these studies, traders' cognitive ability was measured by the cognitive reflection test (CRT Frederick, 2005). The original format is a three-question test. A subject is considered to have a higher cognitive ability if he can solve more of the questions correctly. In all these experiments, the subjects were unable to learn the REE, but the deviation of individual price forecast and market price are much greater in markets populated by participants with lower CRT scores than those populated by participants with higher CRT scores.

Observation 13: Unlike in LtOEs with double auction, the evidence on whether bubbles and crashes can be eliminated when subjects become more experienced is mixed in LtFEs.

Support: One important finding in the experimental literature on LtOEs in which subjects trade in continuous double auction markets is that bubbles tend to disappear when identical markets are repeated and subjects are more experienced with the setting (Dufwenberg et al., 2005; Hussam et al., 2008). In Dufwenberg et al. (2005), the authors made the distinction between *rounds* and *periods*; a round (being a market, or a repetition of the market) consists of ten periods. In the first three rounds, subjects participated in the same double auction market. Though large bubbles and crashes may happen in the first two rounds, the average transaction price of the asset will be very close to the fundamental value in the third round when the subjects become experienced with the experimental setting. In the fourth round, some experienced traders were replaced by inexperienced traders (subjects who did not participate in the experimental market in previous rounds), and the authors found that the asset price remains close to the fundamental value, even when four experienced traders are replaced by inexperienced traders.

To test whether the same result will hold in LtF markets, Kopányi-Peucker and Weber (2021) studied the subjects' expectation formation behavior in LtFE similar to Hommes et al. (2008) for three rounds (markets or repetitions of markets) of 25–40 periods. Their results show that different from in double auction LtOEs, the bubbles and crashes in LtFEs do not disappear in later rounds (when subjects become more experienced). Instead, the bubbles and crashes start even earlier in later rounds than in the first round.

Hennequin (2019) studied if the type of experience matters for its impact on bubble formation with experienced subjects. She ran a two-stage LtFE like Hommes et al. (2008). In Stage 1, each subject is the only human subject in his or her market, and the other 5 subjects are robot players who submit the same price forecast as subjects in a market from a previous experiment. The robots can be Stable Robots who submit forecasts that will lead to a stable convergence to the fundamental value of the asset, or Bubbly Robots who submit forecasts that will lead to persistent bubbles and crashes. In Stage 2, there were three treatments: 6B, where all 6 subjects played with Bubbly Robots in Stage 1; 6S, where all 6 subjects played with Stable Robots in Stage 1; and 3S3B, where 3 subjects played with Stable Robots and 3 subjects played with Bubbly Robots in Stage 1. The result of the experiment shows that the type of experience indeed matters. The asset price is stable in all markets in Treatment 6S, volatile in all markets in Treatment 6B, and may either stabilize or destabilize in Treatment 3S3B.

4. Conclusion

Expectation formation plays a significant role in modern economic modeling of the macroeconomy as well as financial markets. Understanding expectations formation is crucial in designing policies to enhance market stability and manage expectations during a crisis. A LtFE is an experimental methodology aiming at eliciting expectations in the most direct and clean way for the researchers to understand which factors influence market stability via the expectation channel. The result of the LtFE literature usually suggests that while negative feedback markets have a natural tendency to converge to the REE, (strongly) positive feedback markets have a natural and intrinsic tendency to generate bubbles and crashes due to agents coordinating on a common trend in past prices.

By reviewing the result of around 50 recent studies using the LtFE methodology, we show that the findings in the standard LtFE literature are robust against a large variety of changes in the experimental design. The bubbles and crashes found in standard LtFEs with positive expectation feedbacks are also prevalent in experiments with more than 100 subjects in each market. There is high consistency between the findings based on expectation data from LtFEs and survey data from the field. Meanwhile, recent studies also provide a few possible policy tools, e.g., higher interest rate, long horizon forecasts, or higher time pressure, all of which may help manage the trend-following behavior and market oscillations. Individual expectations and market price dynamics in LtFEs are also influenced by the complexity of the task and subjects' cognitive ability, but there is mixed evidence on whether providing subjects full information about the economy will help them learn the REE.

In our views, future work is still needed to address several open questions related to LtFE and expectations formation in macroeconomics and finance in general:

1. Since financial and macroeconomic policy decisions are usually made by professionals like fund managers and central bankers, it may be desirable to run more LtFEs using professional subjects to further strengthen the external

validity of the findings in the literature and to identify the possible differences between students and professional subject pools.

2. There have been studies in which the subjects forecast either exogenous variables (like in Hey, 1994; Afrouzi et al., 2020), variables with expectation feedbacks (standard LtFEs), or real data from the field (Andreassen and Kraus, 1990; De Bondt, 1993), but there is a lack of dialogue among the three strands of literature. Future research may combine the three designs in one experiment and investigate whether there are systematic differences in subjects' forecasting behavior when they predict different types of data.
3. It may be useful to conduct more studies on the relationship between different psychological factors and expectation-formation behavior. For example, how do overconfidence, theory of mind, and emotion influence individual learning and market stability? It may also be interesting to introduce the usage of eye tracking machines and face readers in LtFEs to take more precise measurement of related variables.
4. People usually make point prediction in an economic system in a standard LtFE. In future studies, it may be interesting to study the cases where people make "structural expectations", i.e., providing the confidence interval of a variable, (more than one) parameters in the model instead of the point prediction.
5. Researcher may also try to apply new quantitative or statistical methods to data from LtFEs. For example, instead of running simple regressions, future research may also consider applying the machine-learning method to estimate the coefficients of the heuristics or use artificial intelligence to categorize the subjects into different behavioral types.
6. Though Hommes et al. (2005) already introduced passive, fundamental robot traders in LtFEs. It may be interesting for future LtFEs to introduce more and "smarter" robot traders to determine how human-robot interaction influences individual learning and market stability in LtFEs.

References

- Adam, K., 2007. Experimental evidence on the persistence of output and inflation. *Econ. J.* 117 (520), 603–636.
- Afrouzi, H., Kwon, S.Y., Landier, A., Ma, Y., Thesmar, D., 2020. Overreaction and Working Memory (No. W27947). National Bureau of Economic Research.
- Ahrens, S., Lustenhouwer, J., Tettamanzi, M., 2019. The Stabilizing Effects of Publishing Strategic Central Bank Projections. Technische Universitat Berlin Working Paper.
- Akiyama, E., Hanaki, N., Ishikawa, R., 2017. It is not just confusion! strategic uncertainty in an experimental asset market. *Econ. J.* vol. 127 (605), F563–F580.
- Andreassen, P.B., Kraus, S.J., 1990. Judgmental extrapolation and the salience of change. *J. Forecast.* vol. 9 (4), 347–372.
- Anufriev, M., Arifovic, J., Ledyard, J., Panchenko, V., 2013. Efficiency of continuous double auctions under individual evolutionary learning with full or limited information. *J. Evol. Econ.* vol. 23 (3), 539–573.
- Anufriev, M., Bao, T., Tuinstra, J., 2016. Microfoundations for switching behavior in heterogeneous agent models, An experiment. *J. Econ. Behav. Organ.* vol. 129, 74–99.
- Anufriev, M., Chernulich, A., Tuinstra, J., 2018. A laboratory experiment on the heuristic switching model. *J. Econom. Dynam. Control* vol. 91, 21–42.
- Anufriev, M., Chernulich, A., Tuinstra, J., 2020a. Complexity of Forecasting and Coordination in Experimental Markets with Different Horizons. Mimeo.
- Anufriev, M., Duffy, J., Panchenko, V., 2019a. Planar beauty contests. Available at SSRN 3404334.
- Anufriev, M., Hommes, C., 2012. Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *Am. Econ. J. Microecon.* vol. 4 (4), 35–64.
- Anufriev, M., Hommes, C., Makarewicz, T., 2019b. Simple forecasting heuristics that make us smart: Evidence from different market experiments. *J. Eur. Econom. Assoc.* vol. 17 (5), 1538–1584.
- Anufriev, M., Neunhoeffer, F., Tuinstra, J., 2020b. Pushed To Perform -Time Pressure in Long Run Learning-To-Forecast Experiments. Mimeo.
- Arifovic, J., 1996. The behavior of the exchange rate in the genetic algorithm and experimental economies. *J. Polit. Econ.* vol. 104 (3), 510–541.
- Arifovic, J., 1997. Strategic uncertainty and the genetic algorithm adaptation. In: *Computational Approaches To Economic Problems*. Springer, Boston, MA, pp. 225–236.
- Arifovic, J., Duffy, J., 2018. Heterogeneous agent modelling experimental evidence. In: *Handbook of Computational Economics*, Vol. 4. Elsevier, pp. 491–540.
- Arifovic, J., Hommes, C., Salle, I., 2019. Learning to believe in simple equilibria in a complex OLG economy-evidence from the lab. *J. Econom. Theory* vol. 183, 106–182.
- Arifovic, J., Ledyard, J., 2011. A behavioral model for mechanism design, Individual evolutionary learning. *J. Econ. Behav. Organ.* vol. 78 (3), 374–395.
- Arifovic, J., Ledyard, J., 2012. Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism. *J. Public Econ.* vol. 96 (9–10), 808–823.
- Arifovic, J., Petersen, L., 2017. Stabilizing expectations at the zero lower bound: Experimental evidence. *J. Econom. Dynam. Control* vol. 82, 21–43.
- Assenza, T., Bao, T., Hommes, C., Massaro, D., 2014. Experiments on expectations in macroeconomics and finance. *Exp. Macroecon.* vol. 17 (2014), 11–70.
- Assenza, T., Heemeijer, P., Hommes, C.H., Massaro, D., 2019. Managing self-organization of expectations through monetary policy a macro experiment. *J. Monetary Econ.* 117, 170–186.
- Bao, T., Duffy, J., 2016. Adaptive versus eductive learning: Theory and evidence. *Eur. Econ. Rev.* vol. 83, 64–89.
- Bao, T., Duffy, J., 2021. Signal extraction: experimental evidence. *Theory and Decision* vol. 90 (2), 219–232.
- Bao, T., Duffy, J., Hommes, C., 2013. Learning, forecasting, and optimizing: An experimental study. *Eur. Econ. Rev.* vol. 61, 186–204.
- Bao, T., Duffy, J., Zhu, J., 2019. Asset Pricing with Ambiguous Signals: An Experimental Study. UC Irvine Working Paper.
- Bao, T., Hennequin, M., Hommes, C., Massaro, D., 2020. Coordination on bubbles in large-group asset pricing experiments. *J. Econom. Dynam. Control* vol. 110, 103702.
- Bao, T., Hommes, C., 2019. When speculators meet suppliers positive versus negative feedback in experimental housing markets. *J. Econom. Dynam. Control* vol. 107, 103730.
- Bao, T., Hommes, C., Makarewicz, T., 2017. Bubble formation and (In) efficient markets in learning-to-forecast and optimise experiments. *Econ. J.* vol. 127 (605), F581–F609.
- Bao, T., Hommes, C., Sonnemans, J., Tuinstra, J., 2012. Individual expectations, limited rationality and aggregate outcomes. *J. Econom. Dynam. Control* vol. 36 (8), 1101–1120.
- Bao, T., Zong, J., 2019. The impact of interest rate policy on individual expectations and asset bubbles in experimental markets. *J. Econom. Dynam. Control* vol. 107, 103735.
- Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2020. Overreaction in macroeconomic expectations. *Amer. Econ. Rev.* vol. 110 (9), 2748–2782.
- Bosch-Rosa, C., Meissner, T., Bosch-Domènech, A., 2018. Cognitive bubbles. *Exper. Econ.* vol. 21 (1), 132–153.
- Branch, W., Evans, G.W., McGough, B., 2012. Finite horizon learning. In: Sargent, T., Vilmunen, J. (Eds.), *Macroeconomics At the Service of Public Policy*. Oxford University Press.
- Brock, W.A., Hommes, C.H., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *J. Econom. Dynam. Control* vol. 22 (8–9), 1235–1274.
- Chen, S.H., 2012. Varieties of agents in agent-based computational economics: A historical and an interdisciplinary perspective. *J. Econom. Dynam. Control* vol. 36 (1), 1–25.
- Chen, S.H., 2013. *Evolutionary Computation in Economics and Finance (Part of the Studies in Fuzziness and Soft Computing Book Series, Volume 100)*. Springer Physica-Verlag Heidelberg.
- Chen, S.H., Du, Y.R., Yang, L.X., 2014. Cognitive capacity and cognitive hierarchy, a study based on beauty contest experiments. *J. Econ. Interact. Coord.* vol. 9 (1), 69–105.
- Chen, S.H., Gostoli, U., Tai, C.C., Shih, K.C., 2011. To whom and where the hill becomes difficult to climb, Effects of personality and cognitive capacity in experimental DA markets. *Adv. Behav. Finan. Econ.*
- Chen, S.H., Hsieh, Y.L., 2011. Reinforcement learning in experimental asset markets. *East. Econ. J.* vol. 37 (1), 109–133.
- Cheung, S.L., Hedegaard, M., Palan, S., 2014. To see is to believe: Common expectations in experimental asset markets. *Eur. Econ. Rev.* vol. 66, 84–96.
- Chung, H., Davig, T., Leeper, E.M., 2007. Monetary and fiscal policy switching. *J. Money Credit Bank.* vol. 39 (4), 809–842.
- Cohn, A., Engelmann, J., Fehr, E., Maréchal, M.A., 2015. Evidence for countercyclical risk aversion: An experiment with financial professionals. *Amer. Econ. Rev.* vol. 105 (2), 860–885.
- Coibion, O., Gorodnichenko, Y., 2012. What can survey forecasts tell us about information rigidities? *J. Polit. Econ.* vol. 120 (1), 116–159.

- Coibion, O., Gorodnichenko, Y., 2015. Information rigidity and the expectations formation process: A simple framework and new facts. *Amer. Econ. Rev.* vol. 105 (8), 2644–2678.
- Colasante, A., Alfaro, S., Camacho, E., Gallegati, M., 2018. Long-run expectations in a learning-to-forecast experiment. *Appl. Econ. Lett.* vol. 25 (10), 681–687.
- Colasante, A., Alfaro, S., Camacho-Cuena, E., Gallegati, M., 2020. Long-run expectations in a learning-to-forecast experiment: a simulation approach. *J. Evol. Econ.* vol. 30, 75–116.
- Comand, C., Hubert, P., 2020. On the external validity of experimental inflation forecasts: A comparison with five categories of field expectations. *J. Econom. Dynam. Control* vol. 110, 103746.
- Comand, C., M'baye, C.K., 2018a. Does inflation targeting matter? An experimental investigation. *Macrocon. Dyn.* vol. 22 (2), 362–401.
- Comand, C., M'baye, C.K., 2018b. Band or point inflation targeting? An experimental approach. *J. Econ. Interact. Coord.* vol. 13 (2), 283–309.
- De Bondt, W.P., 1993. Betting on trends: Intuitive forecasts of financial risk and return. *Int. J. Forecast.* vol. 9 (3), 355–371.
- DeGroot, M.H., 2004. *Optimal Statistical Decisions*. John Wiley & Sons Inc., Hoboken.
- Duffy, J., 2006. Agent-based models and human subject experiments. In: *Handbook of Computational Economics*. vol. 2, pp. 949–1011.
- Duffy, J., 2010. Experimental macroeconomics. In: *In Behavioural and Experimental Economics*. Palgrave Macmillan, London, pp. 113–119.
- Duffy, J. (Ed.), 2014. *Experiments in Macroeconomics*. Emerald Group Publishing.
- Duffy, J., Nagel, R., 1997. On the robustness of behaviour in experimental 'beauty contest' games. *Econ. J.* vol. 107 (445), 1684–1700.
- Dufwenberg, M., Lindqvist, T., Moore, E., 2005. Bubbles and experience, an experiment. *Amer. Econ. Rev.* vol. 95 (5), 1731–1737.
- Evans, G.W., Guse, E., Honkapohja, S., 2008. Liquidity traps, learning and stagnation. *Eur. Econ. Rev.* vol. 52 (8), 1438–1463.
- Evans, G.W., Hommes, C.H., McGough, B., Salle, I., 2019. Are long-horizon expectations (de-) stabilizing? In: *Theory and Experiments* (No. 2019-27). Bank of Canada Staff Working Paper.
- Evans, G.W., Honkapohja, S., 1999. Learning dynamics. *Handb. Macrocon.* vol. 1, 449–542.
- Evans, G.W., Honkapohja, S., 2003. Expectations and the stability problem for optimal monetary policies. *Rev. Econom. Stud.* vol. 70 (4), 807–824.
- Evans, G.W., Honkapohja, S., 2009. Learning and macroeconomics. *Annu. Rev. Econ.* vol. 1 (1), 421–449.
- Evans, G.W., Honkapohja, S., 2012. *Learning and Expectations in Macroeconomics*. Princeton University Press.
- Fehr, E., Tyran, J.R., 2005. Individual irrationality and aggregate outcomes. *J. Econ. Perspect.* vol. 19 (4), 43–66.
- Fehr, E., Tyran, J.R., 2008. Limited rationality and strategic interaction, the impact of the strategic environment on nominal inertia. *Econometrica* vol. 76 (2), 353–394.
- Fischbacher, U., Hens, T., Zeisberger, S., 2013. The impact of monetary policy on stock market bubbles and trading behavior: Evidence from the lab. *J. Econom. Dynam. Control* vol. 37 (10), 2104–2122.
- Frederick, S., 2005. Cognitive reflection and decision making. *J. Econ. Perspect.* vol. 19 (4), 25–42.
- Giamattei, M., Huber, J., Lamsdorff, J.G., Nicklisch, A., Palan, S., 2020. Who inflates the bubble? Forecasters and traders in experimental asset markets. *J. Econom. Dynam. Control* vol. 110, 103718.
- Giusti, G., Jiang, J.H., Xu, Y., 2016. Interest on cash, fundamental value process and bubble formation, an experimental study. *J. Behav. Exp. Financ.* vol. 11, 44–51.
- Glaser, M., Iliewa, Z., Weber, M., 2019. Thinking about prices versus thinking about returns in financial markets. *J. Finance* vol. 74 (6), 2997–3039.
- Grandmont, J.M., 1985. On endogenous competitive business cycles. *Econometrica* 995–1045.
- Grosskopf, B., Nagel, R., 2008. The two-person beauty contest. *Games Econom. Behav.* vol. 62 (1), 93–99.
- Guesnerie, R., 1992. An exploration of the eductive justifications of the rational-expectations hypothesis. *Am. Econ. Rev.* vol. 125, 4–1278.
- Hanaki, N., Akiyama, E., Ishikawa, R., 2018. Effects of different ways of incentivizing price forecasts on market dynamics and individual decisions in asset market experiments. *J. Econom. Dynam. Control* vol. 88, 51–69.
- Hanaki, N., Hommes, C., Kopányi, D., Tuinstra, J., 2019a. The Effect of Framing on the Emergence of Bubbles and Crashes in a Learning To Forecast Experiment. Mimeo.
- Hanaki, N., Koriyama, Y., Sutan, A., Willinger, M., 2019b. The strategic environment effect in beauty contest games. *Games Econom. Behav.* vol. 113, 587–610.
- Haruvy, E., Lahav, Y., Noussair, C.N., 2007. Traders' expectations in asset markets, experimental evidence. *Amer. Econ. Rev.* vol. 97 (5), 1901–1920.
- He, S., Kucinskis, S., 2019. Complexity and expectation formation. Available at SSRN 3450207.
- Heinemann, F., Nagel, R., Ockenfels, P., 2004. The theory of global games on test: experimental analysis of coordination games with public and private information. *Econometrica* vol. 72 (5), 1583–1599.
- Hennequin, M., 2019. Experiences and Expectations in Asset Markets, An Experimental Study. Working Paper, University of Amsterdam Working Paper.
- Hennequin, M., Hommes, C., 2019. Managing Bubbles in Experimental Asset Markets with Monetary Policy. Mimeo.
- Hey, J.D., 1994. Expectations formation, rational or adaptive or...? *J. Econ. Behav. Organ.* vol. 25 (3), 329–349.
- Hommes, C.H., 2011. The heterogeneous expectations hypothesis: Some evidence from the lab. *J. Econom. Dynam. Control* vol. 35 (1), 1–24.
- Hommes, C.H., 2013a. Reflexivity, expectations feedback and almost self-fulfilling equilibria: economic theory, empirical evidence and laboratory experiments. *J. Econ. Methodol.* vol. 20, 406–419.
- Hommes, C.H., 2013b. *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*. Cambridge University Press, Cambridge.
- Hommes, C.H., 2014. Behaviourally rational expectations, and almost self-fulfilling equilibria. *Rev. Behav. Econ.* vol. 1, 7597.
- Hommes, C.H., 2021. Behavioral and experimental macroeconomics and policy analysis, a complex systems approach. *J. Econ. Lit.* 59 (1), 149–219.
- Hommes, C., Kopányi-Peuker, A., Sonnemans, J., 2021. Bubbles, crashes and information contagion in large-group asset market experiments. *Exp. Econ.* vol. 24 (2), 414–433.
- Hommes, C., Lux, T., 2013. Individual expectations and aggregate behavior in learning-to-forecast experiments. *Macrocon. Dyn.* vol. 17 (2), 373–401.
- Hommes, C., Makarewicz, T., Massaro, D., Smits, T., 2017. Genetic algorithm learning in a new keynesian macroeconomic setup. *J. Evol. Econ.* vol. 27 (5), 1133–1155.
- Hommes, C., Massaro, D., Salle, I., 2019a. Monetary and fiscal policy design at the zero-lower bound: Evidence from the lab. *Econ. Inq.* vol. 57 (2), 1120–1140.
- Hommes, C., Massaro, D., Weber, M., 2019b. Monetary policy under behavioral expectations: Theory and experiment. *Eur. Econ. Rev.* vol. 118, 193–212.
- Hommes, C., Sonnemans, J., Tuinstra, J., Van de Velden, H., 2005. Coordination of expectations in asset pricing experiments. *Rev. Financ. Stud.* vol. 18 (3), 955–980.
- Hommes, C., Sonnemans, J., Tuinstra, J., Van de Velden, H., 2008. Expectations and bubbles in asset pricing experiments. *J. Econ. Behav. Organ.* vol. 67 (1), 116–133.
- Hommes, C.H., Sonnemans, J., Velden, H., 2000. Expectation formation in a cobweb economy: some one person experiments. In: Delli Gatti, D., Gallegati, M., Kirman, A.P. (Eds.), *Interaction and Market Structure*. Springer Verlag Berlin, pp. 253–266, (2000).
- Hussam, R.N., Porter, D., Smith, V.L., 2008. Thar she blows, can bubbles be rekindled with experienced subjects? *Amer. Econ. Rev.* vol. 98 (3), 924–937.
- Kopányi-Peuker, A., Weber, M., 2021. Experience does not eliminate bubbles: Experimental evidence. *Rev. Financ. Stud.* 34 (9), 4450–4485.
- Kryvtsov, O., Petersen, L., 2013. Expectations and Monetary Policy: Experimental Evidence (No. 2013-44). Bank of Canada Working Paper.
- Kryvtsov, O., Petersen, L., 2021. Central bank communication that works, lessons from lab experiments. *J. Monetary Econ.* 117, 760–780.
- Landier, A., Ma, Y., Thesmar, D., 2019. Biases in expectations: Experimental evidence. Available at SSRN 3046955.
- Lei, V., Noussair, C.N., 2002. An experimental test of an optimal growth model. *Amer. Econ. Rev.* vol. 92 (3), 549–570.
- Li, K.K., 2020. How does the COVID-19 outbreak affect people's expectation about the macroeconomy?. Available at SSRN 3567937.
- Lucas, R.E., 1972. Expectations and the neutrality of money. *J. Econom. Theory* vol. 4, 103–124.
- Lucas Jr., R.E., 1978. Asset prices in an exchange economy. *Econometrica J. Econometric Soc.* 1429–1445.
- Makarewicz, T.A., Anufriev, M., Hommes, C.H., 2020. Evolution of Learning Mechanisms in the Ultimatum Game. Mimeo.
- Mankiw, N.G., Reis, R., 2002. Sticky information versus sticky prices a proposal to replace the new Keynesian Phillips curve. *Q. J. Econ.* vol. 117 (4), 1295–1328.
- Mankiw, N.G., Reis, R., 2007. Sticky information in general equilibrium. *J. Eur. Econom. Assoc.* vol. 5 (2–3), 603–613.
- Marimon, R., Spear, S.E., Sunder, S., 1993. Expectationally driven market volatility: an experimental study. *J. Econom. Theory*.
- Marimon, R., Sunder, S., 1993. Indeterminacy of equilibria in a hyperinflationary world: experimental evidence. *Econometrica* 107, 3–1107.
- Marimon, R., Sunder, S., 1994. Expectations and learning under alternative monetary regimes: an experimental approach. *Econom. Theory* 4 (1), 131–162.
- Marimon, R., Sunder, S., 1995. Does a constant money growth rule help stabilize inflation?: experimental evidence. In: *Carnegie-Rochester Conference Series on Public Policy*, Vol. 43. North-Holland, pp. 111–156.
- Mauersberger, F., 2019. Monetary policy rules in a non-rational world: A macroeconomic experiment. Available at SSRN 3060341.
- Mirdamadi, M., Petersen, L., 2018. Macroeconomic literacy and expectations. Available at SSRN 3187131.
- Muth, J.F., 1961. Rational expectations and the theory of price movements. *Econometrica* 31, 5–335.

- Nagel, R., 1995. Unraveling in guessing games: An experimental study. *Am. Econ. Rev.* 85 (5), 1313–1326.
- Palan, S., 2013. A review of bubbles and crashes in experimental asset markets. *J. Econ. Surv.* 27 (3), 570–588.
- Pfajfar, D., Žakelj, B., 2018. Inflation expectations and monetary policy design: Evidence from the laboratory. *Macroeconomic Dynamics* 22 (4), 1035–1075.
- Smith, V.L., Suchanek, G.L., Williams, A.W., 1988. Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica* 1119–1151.
- Sonnemans, J., Tuinstra, J., 2010. Positive expectations feedback experiments and number guessing games as models of financial markets. *J. Econ. Psychol.* 31 (6), 964–984.
- Stöckl, T., Huber, J., Kirchler, M., 2010. Bubble measures in experimental asset markets. *Exp. Econ.* 13 (3), 284–298.
- Sutan, A., Willinger, M., 2009. Guessing with negative feedback: An experiment. *J. Econom. Dynam. Control* 33 (5), 1123–1133.
- Tai, C.C., Chen, S.H., Yang, L.X., 2018. Cognitive ability and earnings performance: Evidence from double auction market experiments. *J. Econom. Dynam. Control* 91, 409–440.
- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185 (4157), 1124–1131.
- Woodford, M., 2013. Macroeconomic analysis without the rational expectations hypothesis. *Annu. Rev. Econ.* 5 (1), 303–346.
- Zhu, J., Bao, T., Chia, W.M., 2019. Evolutionary selection of forecasting and quantity decision rules in experimental asset markets. *J. Econ. Behav. Organ.* 182, 363–404.
- Zong, J., Fu, J., Bao, T., 2017. Cognitive ability of traders and financial bubbles: An experimental study. *J. World Econ.* 6, 167–192.