## Expectation Formation in Finance and Macroeconomics: A Review of New Experimental Evidence<sup>1</sup>

Te Bao, Cars Hommes, and Jiaoying Pei

**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.

**Keywords**: Learning to Forecast Experiment; Experimental Finance, Rational Expectations; Bubbles and Crashes; Behavioral Finance

JEL Classification: C10, D17, D84, E52, G12, G17, G40

<sup>&</sup>lt;sup>1</sup> Te Bao, School of Social Sciences and NTU-WeBank JRC, Nanyang Technological University Singapore, Email: <u>baote@ntu.edu.sg</u>; Cars Hommes, CeNDEF, School of Economics, University of Amsterdam and Bank of Canada, Email: <u>c.h.hommes@uva.nl</u>; Jiaoying Pei, School of Social Sciences, Nanyang Technological University Singapore. Email: <u>peij0003@e.ntu.edu.sg</u>. Te Bao thanks the financial support from the AcRF Tier 1 Grant RG69/19 from Ministry of Education of Singapore, NTU-WebBank Joint Research Center (NWJ-2019-001, NWJ-2020-003) and the National Science Foundation of China (No. 71803201, 71773013, and 71873149).

#### 1. Introduction

Expectation formation plays a central role in modern finance and macroeconomic modelling. Since the seminal works by Muth (1961) and Lucas (1972), the rational expectations hypothesis (REH) has become the standard approach to model expectation formation. However, due to the lack of highquality 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) and Marimon, Spear, and Sunder (1993) has been widely used by experimental economists to study expectation formation in financial markets and macroeconomies. The key feature of a LtFE is that the subjects of the experiment play the role of professional forecasters (Hommes, 2011, 2013b, 2020). 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 and then determine the realization of the variables on which the subjects made their forecasts. The learning to forecast 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). Because he LtFE design elicits and incentivizes the individual expectations directly, the subjects should have stronger incentives to form rational expectations.

The market can display positive feedback 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, a lower realized market price (Hommes, 2013, 2020). A general conclusion from the LtFE literature is that the agents can learn. The rational expectations equilibrium when the market is a negative feedback system (e.g., Hommes et al., 2000). Yet, agents fail to learn to rational expectation equilibrium when the market is a positive feedback system (e.g., Hommes et al., 2005, 2008). There have been several comprehensive surveys on this literature (e.g., Hommes, 2020, 2011, 2013a, 2013b, 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 the recent trend and discuss the possible future directions of the 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 and the learning to optimize design (e.g., Bao et al., 2013, 2017, Mirdamadi and Petersen, 2018, Giamattei et al., 2020).
- 2) Traditionally, like other market experiments, the market size of a LtFE is 6-10 participants. In recent years, researchers start to run large scale learning to forecast experiments (e.g., Bao, Hennequin, Hommes and Massaro, 2019, Hommes, Kopányi-Peuker and Sonnemans, 2020) to test if the results from relatively small-scale experiments are robust in larger experimental markets.
- 3) A typical LtFE usually lasts for 50 periods, and the predictions are made one period or two periods ahead. In recent years, researchers start to run LtFE with longer horizons and investigate the role of long-run predictions (Colasante et al., 2018; Evans et al., 2019).

- 4) Traditionally, LtFEs on asset market elicits beliefs on asset prices. Recently, some studies compare the cases where agents form expectations on prices versus returns (Glaser et al., 2019, Hanaki et al. (2020)).
- 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, 2019b, Mauersberger, 2019, Ahrens et al., 2020).
- 6) Papers trying 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, 2019).
- 7) Studies that try to compare data on expectation formation from the lab and from the field (e.g., Landier et al., 2019, Cornand and Hubert, 2020).
- 8) Studies on how the complexity of the decision influences the forecasting behavior (e.g., Mirdamadi and Petersen, 2018, Anufriev et al., 2019, Arifovic et al., 2019, Bao and Duffy, 2019, He and Kucinskas, 2019).

In the rest of the paper, we will first go through the basic setup of a learning to forecast experiment in Section 2. After that, we will list the main results that answer the above questions according to the recent literature in Section 3. Finally, we draw a short conclusion and discussion based on the development of the literature in Section 4.

#### 2. Basic Setup of a Learning to Forecast Experiment

#### 2.1 Experimental Design

A baseline learning to forecast experiment is usually a market experiment with6?-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., price of a product/financial asset, inflation rate, GDP output gap, etc. To provide the right 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 learning to forecast experiment, the subjects usually play the role of a professional consult/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 of the firm, and better compensation to the subject.

A learning to forecast experiment 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 the information on others' expectations in the same period either, and even after the realization of the economic variable is revealed. In a way, a learning to forecast experiment 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 learning to forecast experiment is usually considered a price taker, who does not put a lot of consideration on his/her market power in their decision making.

A key research question of the learning to forecast experiment literature is that: when people do not start from the rational expectations equilibrium and do not have the knowledge about the specification of the DGP of the economy, can they learn to play rational expectations over time? Stated differently, can learning lead the market to converge to its rational expectations equilibrium (REE)? According to the rational expectations hypothesis (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. Their prediction errors do not have a cross-sectional correlation.

2.2 Price Dynamics and Individual Expectations



Figure 1: Price dynamics in negative (left panel) and positive feedback (right panel) markets in the learning to forecast experiment by Bao, Hommes, Sonnemans, and Tuinstra (2012).

Figure 1 shows the aggregate price dynamics in a typical learning to forecast experiment. 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.



Figure 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 learning to forecast experiment by Bao, Hommes, Sonnemans, and Tuinstra (2012).

To better understand the individual expectation formation in the experimental markets of learning to forecast experiments, 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 the recent past are assigned with higher evolutionary fitness, and hence attract more followers in the next period. There are typically four forecasting strategies in an

HSM: an adaptive expectations rule, a weak trend following rule (or a contrarian rule), a strong trend extrapolation rule, and an Anchoring and Adjustment rule (Tversky and Kahneman, 1974). As shown by Figure 2, individuals usually follow adaptive or contrarian expectations in negative feedback markets, and strong trend-following rule or anchoring and adjustment rule in positive feedback markets. They are hence 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. Yet, 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) study the price dynamics in experimental housing markets as a "hybrid" of positive and negative feedback systems. The housing market is a production market, and hence a negative feedback system for the builders, and an asset market, and hence 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 other 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.

#### 3. Stylized Results from Recent Literature

In this section, we review the results of some recent studies (most of them 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 try to highlight the main conclusions and supporting evidence, as Palan (2013) did for the literature on asset bubbles in continuous double auction markets a la Smith et al. (1988).

#### 3.1 LtFE versus LtOE

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

**Support**: Since the beginning of the learning to forecast experiment literature, there have been questions about the comparability between the results from LtFEs and learning to optimize experiments (LtOEs), where subjects make quantity decisions directly. Though there have been some LtOEs that also elicit price forecast (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. The underlying model is the same 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. Then the market price is 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 find 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 rational expectations hypothesis works in markets.

Unlike Bao et al. (2013), Bao et al. (2017) study 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 design three treatments: LtFE, LtOE, and a third one called Mixed, where the subject also does both forecasting and quantity decision (on trading) tasks. To avoid potential hedging, the subjects in the Mixed treatment receive their payment based on their performance in the forecasting and trading task with 50:50 probability, instead of 50:50 weight.





Figure 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).

Figure 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 the treatments, the price deviation and the magnitude of fluctuation are way larger in the LtOE and Mixed treatments than in the LtFE treatment.

Besides, the authors 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.

The result from Arifovic et al. (2019) supports the aforementioned fast convergence in LtFE in a complex nonlinear overlapping generations framework. For both the LtFE and LtOE the OLG economy converges to simple equilibria, a steady state, or a 2-cycle. Subjects in LtFE design may converge to a two-cycle, while price predictions in LtOE fail to do so even after the initial oscillations. The authors plot the cumulative distribution of individual decision times and the length of instructions and report 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 lead to the observed differences in outcomes.

Giamattei et al. (2020) find 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 people may wonder if the results of this design are robust when the group size becomes larger. In particular, supporters of the rational expectations hypothesis 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 response to this question, a few recent LtFEs employ large scale design, i.e., by increasing the market size from 6 to 20-30, or even 100. These studies usually show that bubbles and crashes still occur in these large markets as they did in smaller markets.

Bao et al. (2020) study the price dynamics and individual expectations in LtFE markets. 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 increase from 6 to 20-30.



Figure 4: Price Dynamics in Bao et al. (2020).

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

Hommes et al. (2020) extend the size of the large experimental asset market further to around 100 subjects (between 92 to 104) in each market. The unique REE of the asset price in this 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. Besides, the authors also examine the effect of news announcements when the market is overheated and find that it can help to bring down the asset prices substantially.

#### 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) run a learning to forecast experiment where subjects play the role of agents with CRRA utility functions and solve a consumption-based asset pricing problem a la Lucas (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' design four treatments, where the market is populated by 0%, 30%, 50%, and 100% of subjects 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 rational expectations (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 usually users of destabilizing trend following strategies.

Parallel to the paper mentioned above, Anufriev et al. (2020) exam how long-run expectations influence market stability. Different from Evans et al. (2019), their experimental setting is the standard one following 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 introduce 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 stay 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.

Besides these two papers, some studies elicit long-run expectations besides 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. People may 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) run a LtFE where the length increases by a factor three, i.e., to 150 periods. The study shows 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 vary the time pressure faced by the subjects. The subjects have 25 seconds to make their decision in low time pressure treatment, but only 6 seconds in high time pressure treatment. The authors find 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 the prediction is submitted does have an impact on forecasting behavior. Other 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 study finds 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. Others find no significant effect.

**Support**: In the real world, the expectation in financial markets is in terms of returns, and sometimes in price levels. The same applies to the information on the history of these variables. Glaser et al. (2019) study whether the format in which the expectation is elicited and submitted has an impact on expectation formation. They find that, on average, the expectations are higher by between 1.7 and 1.0 percentage points per month if subjects predict returns rather than price levels. In contrast, showing subjects return bar charts as opposed to price line charts leads to a lowered expectation by 1.1 to 2.4 percentage points per month. This finding is robust across subject samples (students versus financial professionals), and whether the payoff for the subjects is fixed or performance-based.

Glaser et al. (2019) is an experiment on expectation formation using an exogenously generated price time series. Thus, 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. (2019) conduct a learning to forecast experiment where the subjects' price or return forecast will be a key variable in determining the asset prices and returns. Hanaki et al. (2019) use 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) is the first experimental study on the impact of monetary policy on asset prices in double auction markets. They find 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 a la Smith et al. (1988). Similarly, Giusti et al. (2016) conclude 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) conduct a LtFE where the REE of the asset price is 60. The initial interest rate for the risk-free asset is 5%. They design 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) and 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.

The result of the experiment shows that the introduction of the monetary policy can reduce 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 by 2/3. 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, instead of the expectation formation process. Figure 5 shows the asset price dynamics in each of the eight markets in Treatment B (top panel) and Treatment P (bottom panel) in Bao and Zong (2019).



Figure 5: The asset price in each of the eight markets in Treatment B (top panel) and Treatment P (bottom panel) in Bao and Zong (2019).

Parallel to Bao and Zong (2019), Hennequin and Hommes (2018) study 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 differentiate 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) test 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 and leading to deviations from the target. Similarly, Kryvtsov and Petersen (2013) also find that Taylor rule monetary policy is a highly effective device in lowering the conditional variance of output gap and inflation. Mauersberger (2019) runs a LtFE in a New Keynesian Economy as in Woodford (2013) and finds that the Taylor principle does not necessarily guarantee convergence to the steady, but the welfare loss due to expectation driven volatility can be largely mitigated by the Taylor principle near the steady-state.

Kryvtsov and Petersen (2020) study if central bank communication can stabilize individual forecasts and aggregate outcomes. They conduct a learning-to-forecast experiment 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, and the subjects know that the interest rate reacts to output and inflation gap, and in a more than one to one to way to 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, all subjects receive an announcement on the central bank's expected policy decision to increase, decrease the interest rate, or let it stay unchanged. Subjects are informed 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 stay unchanged for some periods, and inform the participants in advance on whether the interest rate will be changed or not in the next periods. The authors find 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. (2019)). 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) show 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. (2019) use the nonlinear NK model, while Arifovic and Petersen (2017) use the linear approximation of the model. Arifovic and Petersen introduce autocorrelated shocks to the system, while Hommes et al. (2019) use expectational shocks generated by news announcements.

Despite the differences in the design, the two experiments reach very similar conclusions. They both find that it is very 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. (2020) study 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 as in Assenza et al. (2019), except that they introduce 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 receives two forecasts from the computer: one is a data-driven forecast that is most likely to prevail in the next period (i.e., with the smallest expected prediction error) and the other is a strategic, "required for target" forecast that can help the economy to 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 has to maintain good credibility by publishing projections that are not too far away from realized inflation. In Treatment 3, a computer algorithm publishes 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 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 Rational Expectation Hypothesis, 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 rational expectations in positive feedback markets. Though subjects may learn to play the REE in negative feedback markets, it is also crucial for us 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 Heuristic Switching Model (HSM) based on Brock and Hommes (1997) model, and Genetic Algorithm (GA) models.

The HSM (Anufriev and Hommes, 2012, Hommes et al., 2017, 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) extend the HSM so that it can also be applied to LtOEs.

The more general Genetic Algorithm (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, 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 et al., 2012, Chen, 2013, Hommes and Lux, 2013, Chen, 2014, Hommes et al., 2017, Anufriev et al., 2013, 2019, 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. It 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 has 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 rational expectations paradigm.

**Support**: Cornand and Hubert (2020) carefully collect inflation expectations data from different sources, e.g., experimental data from Pfajfar and Zakelj (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 compare data from the lab and the field and find that they share a high level of common features. All of them 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 Gorodnchenko (2012, 2015).

Landier et al., (2019) compare the expectation formation behavior by the subjects in a forecasting experiment where subjects predict an AR1 time series with the field data from Coibion and Gorodnchenko (2015). They find that in both cases, the rational expectations hypothesis 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) uses an online experiment to elicit subjects' expectations on future growth and inflation in China after the outbreak of COVID-19. He finds 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. They consistently adopt simple belief-formation processes in LtFE regardless of the complexity of the experimental environment, perform worse, and underreact to complicated news, whether it is structured with low or high persistence.

**Support:** Theoretically, 'anything goes' in the equilibrium selection in the complex overlapping generations economy—all equilibria can be supported by a specific learning process. In Arifovic et al. (2019), the authors set up a complex OLG economy and ask which equilibria will be consequently selected and regarded as plausible. Eventually, even in the OLG environment with infinitely many complex periodic and chaotic equilibria, subjects keep using simple belief-formation processes and coordinate on simple equilibria, a steady state, or a 2-cycle. In other words, they keep tracking low-order patterns and considering only recent observations. Indeed, the more sophisticated experimental task requires a higher cognitive load, but in response to the complex environment, subjects' behavioral rules and the aggregate equilibrium may become simpler. As a result, price in almost all experimental economies converges to a neighborhood of a simple perfect-foresight equilibrium, a steady state, or a two-cycle.

He and Kucinskas (2019) study the effect of complexity on expectation formation. Their results show that subjects have difficulty in processing complex information when forming expectations. In their experiments, the subjects need to form expectations on an indicator A which follows AR(1) process. Subjects in the baseline treatment observe only the past realized values of the indicator A. Subjects in the complex treatment additionally observe a leading indicator B that co-generate 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 when they encounter complexity in the complexity treatment, subjects underreact to information on indicator B and exhibit substantially worse performance.

Observation 11: There is mixed 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 learning to forecast experiments, the subjects only 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 learning to forecast experiments with positive feedbacks (e.g., Hommes et al., 2005, 2008), they examine the key factor 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), (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, and providing full information in LtFEs does not help much for subjects to learn to the REE price.

Mirdamadi and Petersen (2018) run 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 vary the level of the information the subjects receive on the data generation process of the economy. They find that providing precise quantitative training can help to reduce the inflation forecast errors and disagreements about inflation and encourages a more substantial reaction to past forecast errors.

Multi-dimensionality is a common form of complexity faced by agents in macroeconomic or finance models. Anufriev, Duffy, and Panchenko (2019) extend 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., 2019) that provides subjects with full information about the DGP for the two endogenous variables. In particular, they focus on the simplest possible two-dimensional structure: variable a is decoupled (that only depends on the average forecast for  $\alpha$  as in a standard beauty contest game), variable b depends on the average predictions of both a and b. The results show that like in univariate LtFEs, while negative feedback markets tend to convergence to the REE, positive feedback markets usually show persistent deviation from the REE also in the planar LtFE. With the full information of DGP, convergence is achieved in the Sink treatment (when the eigenvalues are both less than 1 in absolute value); convergence is also observed in the negative-feedback Saddle treatment—where the sign of the unstable eigenvalue is negative—implying a remarkable convergence by subjects who were not initially being placed on the saddle path, which contradicts the standard economic theory (Ellison and Pearlman, 2011; Evans and Honkapohja, 2012). In contrast, subjects dealing with the positive unstable root in the saddle-path treatment—converge only towards the boundary solution of the system—while simply diverge away from the interior steady-state when absent the boundary.

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 was not tested directly. Bao and Duffy (2019), and Bao et al. (2020) test the theoretical prediction in the laboratory. The subjects play the role of a financial advisor of an investment company. The experiment employs a call market design. The company will be a buyer/seller of the asset if the subjects' prediction is explained by the signal extraction model very well, though there is large heterogeneity in individual expectations. Besides, 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) show that cognitive ability plays an important role in determining whether individuals can form rational expectations. Akiyama et al. (2017) and Bosch-Rosa et al. (2018) employ call market design, and traders' expectation is elicited for the first period only while Zong et al. (2017) employ the LtFE design. In all these studies, traders' cognitive ability is 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 double auction markets, bubbles and crashes in LtFE are not eliminated when subjects become more experienced.*

**Support**: One important finding in the experimental literature on continuous double auction is that bubbles tend to disappear when identical markets are repeated (Haruvy et al., 2007). That is, if subjects participate in the same double auction market for three rounds of 10-20 periods, 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 the third round.

To test whether the same result will hold in learning to forecast markets, Kopányi-Peuker and Weber (2020) study the subjects' expectation formation behavior in learning to forecast experiment similar to Hommes et al. (2008) for three rounds of 25-40 periods. They also vary the amount of information received by the subjects in three treatments: NO\_INFO in which subjects do not know the fundamental value of the asset as in a typical LtFE; INFO\_AFTER in which subjects are informed about the fundamental value of the asset at the end of each round: FULL\_INFO in which subjects are informed the fundamental value of the asset already in the instructions. Their results show that different from in double auction markets, the bubbles and crashes in LtFEs do not disappear in later rounds. Instead, the bubbles and crashes occur in earlier periods in later rounds than in the first round.

Besides, Hennequin (2019) study if the type of experience matters for its impact on bubble formation with experienced subjects. She runs a two-stage learning to forecast experiment 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 are three treatments: 6B, where all 6 subjects played with Bubbly Robots in Stage 1; 6 S 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 very stable in all markets in Treatment 6S, very volatile in all markets in Treatment 6B and may either stabilize or destabilize in Treatment 3S3B.

#### Conclusion

Expectation formation plays a central 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 learning to forecast experiment 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. Meanwhile, recent studies also provide a few possible policy tools, e.g., higher interest rate, longer market horizon, or higher time pressure that may help to manage the trend following behavior and market oscillations. Based on the above processes in the recent literature, we expect that there will be more studies in this field that keep generating new findings and insights on expectation formation as an important individual cognitive process and a determinant of aggregate market outcomes and stability. More lab experiments will help policymakers to learn how to manage the coordination of individual expectations in macro-financial settings.

#### Reference

Adam, K. (2007). Experimental evidence on the persistence of output and inflation. The Economic Journal, 117(520), 603-636.

Ahrens, S., Lustenhouwer, J., & Tettamanzi, M. 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. The Economic Journal, 127(605), F563-F580.

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

Anufriev, M., Arifovic, J., Ledyard, J., & Panchenko, V. (2013). Efficiency of continuous double auctions under individual evolutionary learning with full or limited information. Journal of Evolutionary Economics, 23(3), 539-573.

Anufriev, M., Bao, T., & Tuinstra, J. (2016). Microfoundations for switching behavior in heterogeneous agent models: An experiment. Journal of Economic Behavior & Organization, 129, 74-99.

Anufriev, M., Chernulich, A., & Tuinstra, J. (2018). A laboratory experiment on the heuristic switching model. Journal of Economic Dynamics and Control, 91, 21-42.

Anufriev, M., Duffy, J., & Panchenko, V. (2019). Planar Beauty Contests. Available at SSRN 3404334.

Anufriev, M., Hommes, C., & Makarewicz, T. (2019). Simple forecasting heuristics that make us smart: Evidence from different market experiments. Journal of the European Economic Association, 17(5), 1538-1584.

Anufriev, M., Chernulich, A. & Tuinstra, J. (2020a). Complexity of forecasting and coordination in experimental markets with different horizons. Mimeo.

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. Journal of Political Economy, 104(3), 510-541.

Arifovic, J. (1997). Strategic uncertainty and the genetic algorithm adaptation. In Computational Approaches to Economic Problems (pp. 225-236). Springer, Boston, MA.

Arifovic, J., & Duffy, J. (2018). Heterogeneous agent modelling: experimental evidence. In Handbook of Computational Economics (Vol. 4, pp. 491-540). Elsevier.

Arifovic, J., Hommes, C., & Salle, I. (2019). Learning to believe in simple equilibria in a complex OLG economy-evidence from the lab. Journal of Economic Theory, 183, 106-182.

Arifovic, J., & Ledyard, J. (2011). A behavioral model for mechanism design: Individual evolutionary learning. Journal of Economic Behavior & Organization, 78(3), 374-395.

Arifovic, J., & Ledyard, J. (2012). Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism. Journal of Public Economics, 96(9-10), 808-823.

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

Assenza, T., Bao, T., Hommes, C., & Massaro, D. (2014). Experiments on expectations in macroeconomics and finance. Experiments in macroeconomics, 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. Journal of Monetary Economics, forthcoming.

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

Bao, T., & Duffy, J. (2019). Signal Extraction in the Laboratory. UC Irvine Working Paper.

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. Journal of Economic Dynamics and Control, 110, 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, 103730.

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.

Bao, T., Hommes, C., & Makarewicz, T. (2017). Bubble Formation and (In) Efficient Markets in Learning-to-forecast and optimise Experiments. The Economic Journal, 127(605), F581-F609.

Bao, T., & Zong, J. (2019). The impact of interest rate policy on individual expectations and asset bubbles in experimental markets. Journal of Economic Dynamics and Control, 107, 103735.

Bosch-Rosa, C., Meissner, T., & Bosch-Domènech, A. (2018). Cognitive bubbles. Experimental Economics, 21(1), 132-153.

Branch, W., Evans, G. W. & McGough, B. (2012), Finite Horizon Learning, in T. Sargent & J. Vilmunen, eds, 'Macroeconomics at the Service of Public Policy', Oxford University Press.

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. Advances in Behavioral Finance and Economics.

Chen, S. H. (2012). Varieties of agents in agent-based computational economics: A historical and an interdisciplinary perspective. Journal of Economic Dynamics and Control, 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. Journal of Economic Interaction and Coordination, 9(1), 69-105.

Chen, S. H., & Hsieh, Y. L. (2011). Reinforcement learning in experimental asset markets. Eastern Economic Journal, 37(1), 109-133.

Cheung, S. L., Hedegaard, M., & Palan, S. (2014). To see is to believe: Common expectations in experimental asset markets. European Economic Review, 66, 84-96.

Chung, H., Davig, T., & Leeper, E. M. (2007). Monetary and fiscal policy switching. Journal of Money, Credit and Banking, 39(4), 809-842.

Cohn, A., Engelmann, J., Fehr, E., & Maréchal, M. A. (2015). Evidence for countercyclical risk aversion: An experiment with financial professionals. American Economic Review, 105(2), 860-85.

Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? Journal of Political Economy, 120(1), 116-159.

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. American Economic Review, 105(8), 2644-78.

Colasante, A., Alfarano, S., Camacho, E., & Gallegati, M. (2018). Long-run expectations in a learning-to-forecast experiment. Applied Economics Letters, 25(10), 681-687.

Colasante, 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, 75–116.

Cornand, C., & Hubert, P. (2020). On the external validity of experimental inflation forecasts: A comparison with five categories of field expectations. Journal of Economic Dynamics and Control, 110, 103746.

DeGroot, M.H. (2004). Optimal Statistical Decisions. Hoboken: John Wiley & Sons Inc.

Duffy, J. (2006). Agent-based models and human subject experiments. Handbook of Computational Economics, 2, 949-1011.

Duffy, J. (2010). Experimental macroeconomics. In Behavioural and Experimental Economics (pp. 113-119). Palgrave Macmillan, London.

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. The Economic Journal, 107(445), 1684-1700.

Ellison, M., & Pearlman, J. (2011). Saddlepath learning. Journal of Economic Theory, 146(4), 1500.

Evans, G. W., & Honkapohja, S. (2012). Learning and expectations in macroeconomics. Princeton University Press.

Evans, G. W., Hommes, C. H., McGough, B., & Salle, I. (2019). Are long-horizon expectations (de-) stabilizing? Theory and experiments (No. 2019-27). Bank of Canada Staff Working Paper.

Evans, G. W., & Honkapohja, S. (1999). Learning dynamics. Handbook of Macroeconomics, 1, 449-542.

Evans, G. W., & Honkapohja, S. (2003). Expectations and the stability problem for optimal monetary policies. The Review of Economic Studies, 70(4), 807-824.

Evans, G. W., Guse, E., & Honkapohja, S. (2008). Liquidity traps, learning and stagnation. European Economic Review, 52(8), 1438-1463.

Evans, G. W., & Honkapohja, S. (2009). Learning and macroeconomics. Annual Review of Economics, 1(1), 421-449.

Fehr, E., & Tyran, J. R. (2005). Individual irrationality and aggregate outcomes. Journal of Economic Perspectives, 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, 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. Journal of Economic Dynamics and Control 37 (10), 2104–2122.

Frederick, S. (2005). Cognitive reflection and decision making. Journal of Economic Perspectives, 19(4), 25-42.

Gabaix, X. (2019). Behavioral inattention. In Handbook of Behavioral Economics: Applications and Foundations 1 (Vol. 2, pp. 261-343). North-Holland.

Giamattei, M., Huber, J., Lambsdorff, J. G., Nicklisch, A., & Palan, S. (2020). Who inflates the bubble? Forecasters and traders in experimental asset markets. Journal of Economic Dynamics and Control, 110, 103718.

Giusti, G., Jiang, J.H., Xu, Y., 2016. Interest on cash, fundamental value process and bubble formation: an experimental study. Journal of Behavioral and Experimental Finance, 11, 44–51.

Glaser, M., Iliewa, Z., & Weber, M. (2019). Thinking about prices versus thinking about returns in financial markets. The Journal of Finance, 74(6), 2997-3039.

Grosskopf, B., & Nagel, R. (2008). The two-person beauty contest. Games and Economic Behavior, 62(1), 93-99.

Hanaki, N., Akiyama, E., & Ishikawa, R. (2016). A methodological note on eliciting price forecasts in asset market experiments. Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), University of Nice Sophia Antipolis, Working Papers, 2.

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. Journal of Economic Dynamics and Control, 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 and Economic Behavior, 113, 587-610.

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

He, S. and S. Kucinskas (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, 72(5), 1583-1599.

Hennequin, M., Hommes, C.H., 2018. Managing Coordination on Bubbles in Experimental Asset Markets with Monetary Policy. Mimeo.

Hennequin, M. (2019). Experiences and expectations in asset markets: an experimental study. Working Paper, University of Amsterdam Working Paper.

Hommes, C.H., 2011. The heterogeneous expectations hypothesis: some evidence for the lab. Journal of Economic Dynamics & Control 35, 1-24.

Hommes, C.H., (2013a), Reflexivity, expectations feedback and almost self-fulfilling equilibria: economic theory, empirical evidence and laboratory experiments, Journal of Economic Methodology 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, Review of Behavioral Economics 1, 7597.

Hommes, C.H. (2020), Behavioral and Experimental Macroeconomics and Policy Analysis: A Complex Systems Approach, Journal of Economic Literature, forthcoming.

Hommes, C., & Lux, T. (2013). Individual expectations and aggregate behavior in learning-to-forecast experiments. Macroeconomic Dynamics, 17(2), 373-401.

Hommes, C. H., Kopányi-Peuker, A., & Sonnemans, J. (2019). Bubbles, crashes, and information contagion in large-group asset market experiments.

Hommes, C.H., Sonnemans, J.H., Tuinstra, J. and van de Velden, H, "Coordination of Expectations in Asset Pricing Experiments." Review of Financial Studies, 18 (3), 2005, 955-980.

Hommes, C.H., Sonnemans, J.H., Tuinstra, J., and van de Velden, H., "Expectations and Bubbles in Asset Pricing Experiments." Journal of Economic Behavior and Organization 67, 2008, 116-133.

Hommes, C.H., Sonnemans, J. and Velden, H. van de, (2000), Expectation Formation in a Cobweb Economy: some one Person Experiments, In: Delli Gatti, D., Gallegati, M., and Kirman, A.P. (Eds.) Interaction and Market Structure, Springer Verlag Berlin (2000), pp.253-266.

Hommes, C., Makarewicz, T., Massaro, D., & Smits, T. (2017). Genetic algorithm learning in a New Keynesian macroeconomic setup. Journal of Evolutionary Economics, 27(5), 1133-1155.

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., Massaro, D., & Weber, M. (2019). Monetary policy under behavioral expectations: Theory and experiment. European Economic Review, 118, 193-212.

Kopányi-Peuker, A., & Weber, M. (2020). Experience does not eliminate bubbles: Experimental evidence. Review of Financial Studies, forthcoming.

Kryvtsov, O., & Petersen, L. (2013). Expectations and monetary policy: experimental evidence (No. 2013-44). Bank of Canada Working Paper.

Kryvtsov, O., & Petersen, L. (2020). Central bank communication that works: Lessons from lab experiments. Journal of Monetary Economics, forthcoming.

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. American Economic Review, 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., 1972. Expectations and the neutrality of money. Journal of Economic Theory 4, 103-124.

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. The Quarterly Journal of Economics, 117(4), 1295-1328.

Mankiw, N. G., & Reis, R. (2007). Sticky information in general equilibrium. Journal of the European Economic Association, 5(2-3), 603-613.

Marimon, R., S.E. Spear, and S. Sunder, 1993, "Expectationally Driven Market Volatility: An Experimental Study," Journal of Economic Theory, 61, 74-103.

Marimon, R. and S. Sunder, "Indeterminacy of Equilibria in a Hyperinflationary World: Experimental Evidence," Econometrica, 61(5), 1993, 1073-1107.

Marimon, R. and S. Sunder, "Expectations and Learning under Alternative Monetary Regimes: An Experimental Approach," Economic Theory, 4, 1994, 131-62.

Marimon, R. and S. Sunder, 1995, "Does a constant money growth rule help stabilize inflation?" Carnegie-Rochester Conference Series on Public Policy, 43, 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.E., 1961. Rational expectations and the theory of price movements. Econometrica 29, 315-335.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. The American Economic Review, 85(5), 1313-1326.

Palan, S. (2013). A review of bubbles and crashes in experimental asset markets. Journal of Economic Surveys, 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.

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.

Tai, C. C., Chen, S. H., & Yang, L. X. (2018). Cognitive ability and earnings performance: Evidence from double auction market experiments. Journal of Economic Dynamics and 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. Annual Review of Economics., 5(1), 303-346.

Zhu, J., Bao, T., & Chia, W. M. (2019). Evolutionary selection of forecasting and quantity decision rules in experimental asset markets. Journal of Economic Behavior & Organization. forthcoming.

Zong, J., Fu, J., & Bao, T. (2017). Cognitive Ability of Traders and Financial Bubbles: An Experimental Study. The Journal of World Economy, 6, 167-192.