New Experimental Evidence on Expectations Formation *

Augustin Landier† Yueran Ma‡

David Thesmar§

September 28, 2017

Abstract

In this paper, we measure belief formation in an experimental setting where agents are incentivized to provide accurate forecasts of a random variable, drawn from a stable and simple statistical process. Using these data, we estimate an empirical model that builds on the recent literature on expectation dynamics: It nests rational expectations, but also allows for extrapolation and under-reaction. Our findings are threefold. First, the rational expectation hypothesis is strongly rejected in our setting, and we find little evidence or learning. Second, both extrapolation and underreaction patterns are statistically discernible in the data, but extrapolation quantitatively dominates. Third, our model coefficients are very robust to changes in experimental setting: They do not depend on process parameters, individual characteristics or framing. These large and stable deviations from rationality occur even though the forecasting exercise is simple and transparent.

*We thank David Norris for very skillful research assistance on this project. Landier acknowledges financial support from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013) Grant Agreement no. 312503 SolSys.

†HEC Paris

‡Harvard University

§MIT-Sloan and CEPR

1 Introduction

The way agents update their expectations about future outcomes is at the very core of most economic models. When updating their beliefs, rational agents are supposed to combine new information with their priors using Bayes' rule. By contrast, non-rational agents might either over-react or under-react to new information, leading to predictable forecast errors.

The finance literature is somewhat divided over which effect dominates. A first branch emphasizes *over-reaction*. Very early on, Shiller (1981) observes that stock prices are more volatile than dividends, and explains it via extrapolative expectations: Agents tend to assume that recent trends will continue so that prices move to much in response to recent shifts in fundamental. As a result, good news lead to over-optimistic expectations and forecast lower realized returns. This effect has been invoked to explain fluctuations in stock and bond returns, as well as phenomena such as the value premium and even overinvestment.[1] A second branch of the empirical literature emphasizes the role *under-reaction* in expectation formation. In these papers, good news about fundamental are only slowly incorporated into expectations, so that good news predict positive future returns. Such under-reaction is invoked in the literature to explain momentum within and across stocks, as well as other anomalies such as the post-announcement drift, the repurchase or profitability anomalies, or even the forward premium puzzle.[2]

[1] For instance, Lakonishok et al. (1994) and Laporta (1996) argue that the value premium is related to extrapolative bias. De Bondt (1993) and Greenwood and Shleifer (2014) also find evidence of extrapolation in stock-prices forecasts. Using a household survey on forecasted earnings, (Dominitz, 1998) find that revisions of expectations are positively related to changes in realized individual earnings. More recently, Gennaioli et al. (2015) find that errors in CFO expectations of earnings growth are not rational and result from extrapolative expectations. Bordalo et al. (2017b) show that during booming bond markets (low credit spreads), agents make over-optimistic forecasts (they expect the spreads to remain low). Bordalo et al. (2017a) finds that high expected long-term growth forecasts negative stock returns.

[2] Abarbanell and Bernard (1992) show that analysts under-react to past earnings in their forecasts, which can explain the profitability anomaly (Bouchaud et al. (2016)). Ball and Brown (1968) show that firms experiencing high earnings surprises experience positive abnormal returns going forward: This post-earnings announcement drift suggests that investors under-react to the information content of earnings. Hong et al. (2000); Hou (2007), among others, explain momentum by the excessively slow diffusion of public information into prices. The market tends to under-react to public announcements such as share repurchases announcements (Ikenberry et al. (1995)) or insiders' trades (Lakonishok and Lee (2001)). Cohen and Lou (2012) find that returns of firms that operate in several industries are predictable as the market under-reacts to industry-news regarding these complicated firms. Frankel and Froot (1985) find evidence in under-reaction in expert forecasts of exchange rates. Gourinchas and

The goal of this paper is to directly measure belief formation in an experimental setting where agents are incentivized to provide accurate forecasts of a random variable, drawn from a stable and simple statistical process (an AR1 process). There is a large literature analyzing expectation formation from field data, with a recently renewed interest in the topic (see literature review below). Existing studies find evidence of both under-reaction and over-reaction depending on the economic variable. Our aim here is to complement this literature with evidence from the lab. While we are aware of potential external validity concerns, relying on an experiment has three main advantages over field data. First, we are able to define the process to be forecasted, so that we can overcome the problem that, in non-experimental data, the underlying data generating process is unknown to both the econometrician and the forecaster. This problem makes it difficult to precisely pin down rational vs. irrational updating, as rational agents might for instance assign a probability that the data generating process can change, leading to complex and hard-to-test bayesian updating. In the context of our empirical framework, we inform agents that the data generating process is stable and endow them with enough observations to estimate this process. In short, there is no ambiguity about what is the "rational expectation" in our experiment. The second advantage of our setting is that it is a pure exercise in time series forecasting, where agents' incentives are clearly defined: they just need to make accurate forecasts of a simple process. Also, the process is not polluted by other economic considerations such as strategic considerations or career concerns. The third advantage of running an experiment is more classical: it allows us to control the environment in a fully random way. For instance, we can change the parameters of the stochastic process. We can also modify at will the framing of the experiment. This allows us to fully control the determinants of the expectation formation process, and how robust our findings are to the environment.

By running our experiment, we generate a large panel of participant expectations and realizations of random processes, under different conditions. We use this panel to estimate an expectation

Tornell (2004) calibrate a model where the foreign exchange forward-premium puzzle can arise from investors under-reaction to interest rate shocks.

formation model that includes both under-reaction and extrapolation, but nests rational expectations as a particular case. Under-reaction and extrapolation can be separately identified in the data via the term structure of expectations. Our findings are the following. First, the rational expectation hypothesis is strongly rejected in our setting (this is consistent with earlier experimental studies). This is true both for a majority of participants taken individually, but also for the average individual. On average, the score is approximately 30% below the score expected for a rational forecaster. In addition, we find little evidence that subjective forecasts converge to rational ones after 40 rounds of testing. Second, patterns of extrapolation largely dominate in explaining expectation dynamics. This is in spite of the fact that the forecasting problem is made as simple as possible. Both extrapolation and underreaction patterns are statistically discernible in the data, but extrapolation quantitatively dominates. Our model explains average expectations very well (with an R^2 of 50-60% depending on specifications). Interestingly, both biases to not decline over time, so that learning does not seem to affect expectation biases. Third, our model coefficients are surprisingly robust to experimental setting. They do not depend on the parameters of the model (our conditions span persistence parameters from 0 to 0.8 and various levels of volatility). They do not depend on the way we label the process (GDP, inflation etc.). The extent of under-reaction is a bit affected by the way we ask participant to report the term structure of their expectation, but in all cases, the amplitude of over-reaction is surprisingly stable.

The next Section (Section 2) is devoted to a detailed review of the experimental literature. Section 3 describes the econometric framework that we use. This framework contains a model of belief formation that nests rational, extrapolative and sticky expectations. Section 4 describes the experimental design. Section 5 describes the results. Section 6 concludes.

2 Related literature

In analyzing belief dynamics in the lab, we contribute to the empirical literature on expectation formation in experiments (see for instance Assenza et al. (2014), for a survey of this literature).

Kahneman and Tversky (1973) offer one of the first experimental studies studying biases in forecasts. They show that subjects confuse the likelihood of a certain assertion being true with representativeness of the situation being described.

This experimental literature has taken different routes: Some papers attempt to categorize people into various types of forecasters. Only some of them can be considered rational, while others are for instance adaptive or extrapolative. Using surveys on future stock returns expectations Dominitz and Manski (2011) find that individuals form beliefs according to processes that are heterogeneous across individuals but stable at the individual level. They attempt to classify people into three types of expectation formation families: random-walk, persistence, and mean-reversion. However, they show that this representation of heterogeneity matches the data relatively poorly. Other papers embrace the view that beliefs dynamics can be explained by regime-switches, where subjects think that very different data generating processes are plausible and constantly update their beliefs on which process is more likely. Bloomfield and Hales (2002) run a trading experiment on MBA students and document that participants under-react more to changes when they follow many reversals. They interpret the result as evidence for regime-switching in beliefs dynamics, a la Barberis et al. (1998). Schmalensee (1976) finds evidence that the speed of adjustment of forecasts falls during turning point periods where the data generating process seems to be changing. Last, some of the literature focuses on equilibrium effects in set-ups where the realized variable depends on forecasts (see Hommes (2011)).

Broadly speaking the literature tends to reject simple forms of the rational expectation hypothesis (there are exceptions: In an experience run on 40 subjects, Dwyer et al. (1993b) fails to reject the rational anticipation model) and emphasizes the importance of recent lags in beliefs formation. For instance, Hey (1994b), runs an experiment where a group of 48 undergraduate students are asked to predict future realizations of a time-series drawn from a stable auto-regressive process. The study rejects rational expectations and finds evidence that adaptive expectations and extrapolation have explanatory power on beliefs dynamics. Beshears et al. (2013) shows in a experiment that agents fail to integrate long-term mean-reversion in their forecasts, while they are sensitive

to short-term momentum and short-term mean-reversion. This leads to a form of extrapolative bias whereby recent trends are assumed to last excessively longer. Such long-term extrapolative forecasting is generated theoretically in Fuster et al. (2010): In their model, agents form expectations by estimating growth regressions with a small number of lagged variables whereas the true data generating process has hump-shaped dynamics.

Our paper contributes to the literature by showing the co-existence of both under-reaction and extrapolation at the individual level: We nest both effects in a simple model and find, in our estimation, parameters that are very stable across groups and speed of mean-reversion. We find that agents exhibit extrapolative behavior, but also, at the same time, excess stickiness in their forecasts. For the parameters of the model, we find that extrapolation is the dominating force. We also find that biases do not fade away over time: in our experiment agents do not learn from their past mistakes. A second differentiating feature of our paper is the use of the term-structure of forecasts: we ask agents to make predictions at several horizons. This makes the identification of under-reaction separately from over-reaction more credible. Under-reaction is measured by the extent to which current expectations are "stick to" previous forecasts. Extrapolation is identified off of the fact that current news have too much impact (i.e. beyond what is rational) on forecasts.

3 Expectations: Extrapolative vs. sticky

This Section explicits our econometric model, which nests rational, extrapolating and sticky expectations. Consider a random variable x_t that an agent is trying to forecast. We note $F_{t-k}x_{t+1}$ the subjective forecast of x_{t+1}, k periods ahead of date t. It may differ from $E_{t-k}x_{t+1}$, the full information rational expectation (we run robustness checks using least square learning as we discuss in Section 5.1).

We will write the subjective forecast $F_{t-k}x_{t+1}$ as the sum of an extrapolative (for overreaction) and a sticky (for underreaction) term. Let us first describe how the two parts are defined. Then, we will combine them in a single specification.

We model extrapolative expectations F^e as:

$$F^e_{t-k} x_{t+1} = E_{t-k} x_{t+1} + \gamma (x_{t-k} - E_{t-k-1} x_{t-k}) \tag{1}$$

where γ captures the strength of overreaction. This specification is similar to Bordalo et al. (2017b) and Bordalo et al. (2017a). Extrapolative individuals react too much to unexpected innovation ($\gamma > 0$). If, however, x_t has a deterministic trend and does not deviate from it, subjective expectations will be rational. Hence, only unexpected positive deviation from the trend will generate overoptimistic expectations. Another nice property of this specification is that it nests rational expectations as a special case ($\gamma = 0$).

We model sticky expectations F^s using the recursive formulation:

$$F^s_{t-k} x_{t+1} = (1 - \lambda) E_{t-k} x_{t+1} + \lambda F^s_{t--k1} x_{t+1} \tag{2}$$

where $\lambda \in [0; 1]$ measures the degree of stickiness. $\lambda = 0$ corresponds to fully rational expectations. In this specification (which for $\lambda = 0$ yields rational expectations), the agent is simply lagging in her updates. Sticky expectations can thus be seen as a form of under-reaction as the agent only partially takes into account new informations and remains stuck with forecasts that were rational in the past. This modeling of sticky expectations appears in Coibion and Gorodnichenko (2015) ; It is used by them in a somewhat different context, as they try to model rational stickiness due to limited information, whereas we study irrational stickiness in a world where all agents have the same information. Our specification of sticky expectations is in line with the limited attention literature, where agents do not update their beliefs and consumption plans in continuous time. For instance, in Mankiw and Reis (2001), firms update their pricing plans with Poisson probability λ and Gabaix and Laibson (2001) have a model where agents update their consumption plans periodically instead of continuously. Our modeling of sticky expectations gives a central role to the term-structure of forecasts, which is well suited for our experimental design where agents

provide forecasts at various horizons.

The empirical specification that we use in this paper combines the two formulations in the following manner:

$$F_{t-k}x_{t+1} - E_{t-k}x_{t+1} = \underbrace{\lambda\left(F_{t-k-1}x_{t+1} - E_{t-k}x_{t+1}\right)}_{\text{underreaction}} + \underbrace{\gamma(x_{t-k} - E_{t-k-1}x_{t-k})}_{\text{extrapolation}} \tag{3}$$

In this specification, the individual forecaster can be both extrapolative $\gamma > 0$ and sticky $\lambda > 0$. These two effects can be estimated by regressing the expectation error $F_t x_{t+1} - x_{t+1}$ on past period innovation $x_t - E_{t-1}x_t$, which captures extrapolation, and past period expectation mistake $F_{t-1}x_{t+1} - E_t x_{t+1}$, which captures underreaction. Intuitively, if expectation errors can be forecasted using previous period errors, this is a sign of underreaction, and $\lambda > 0$. If expectation errors can be forecasted using past innovation, this is a sign of extrapolation, and $\gamma > 0$. Hence, individuals can be both under- and over-reacting to news, and the two effects are separately identified from the term structure of subjective forecasts $F_{t-1}x_{t+1}, F_t x_{t+1}$, and rational expectations $E_{t-1}x_{t+1}, E_t x_{t+1}$.

4 Experiment Design

We recruit participants using standard MTurk HITs titled "Making Statistical Forecasts." Participants are adults from across the US. They complete the experiment using their own electronic devices (e.g. computers and tablets). The MTurk platform is commonly used in experimental studies (Kuziemko et al., 2015; D'Acunto, 2015; Cavallo et al., 2016; Dellavigna and Pope, 2017). It offers a large subject pool and a more diverse sample compared to lab experiments. Prior research also finds the response quality on MTurk to be similar to other samples and to lab experiments (Lian et al., 2017; Casler et al., 2013).

4.1 Experimental conditions

Participants first read a consent form (shown in the Survey Appendix), with a brief description of the experiment, the payments, and the duration (described in detail below). Once participants agree to the consent form, they read instructions and start the experiment. In all conditions, they are first presented with 40 historical realizations a statistical process, and are asked to predict future realizations for 40 rounds. After the prediction task, participants answer some basic demographic questions. The specifics of the prediction task vary from condition to condition, and are described in the following paragraphs. We conducted 3 different experiments sequentially (the baseline, and then versions of it). For each experiment, we made sure the participants were different by excluding participants of previous batches.

Experiment 1. Experiment 1 was conducted in February 2017 and is our baseline test. The various conditions are summarized in Table 1, Panel A. In the experiment, each participant is presented with a different realization of an AR1 process:

$$x_{t+1} = \rho x_t + \epsilon_t \tag{4}$$

In each round, participants are asked to predict the value of the next two realizations x_{t+1} and x_{t+2}. Figure 1 provides a screen shot of the prediction page. Specifically, a series of green dots show past realizations of the statistical process. Participants can drag the mouse to indicate their prediction for the next realization, $F_t x_{t+1}$, in the purple bar, and indicate the following realization, $F_t x_{t+2}$, in the red bar. Participants' predictions are shown as yellow dots. We also display the prediction of x_{t+1} from the previous round $F_{t-1} x_{t+1}$ using a grey dot (participants can see it but cannot change it). After making their decisions, participants click "Make Predictions" and move on to the next round.

In this experiment, we use 6 different values of ρ: $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The volatility of ϵ is 20. Each participants is randomly assigned to one value of ρ. Each participant is presented with a

different realization of the process. There are 270 participants in total and about 30 participants per value of ρ (the randomization is not perfectly even across conditions in a finite sample).

Experiment 2. Experiment 2 was conducted in March 2017. Its goal is the study of potential heterogeneity in participants' responses to the same statistical process. Thus we perform an experiment that is similar to Experiment 1, except we use the same value of $\rho = 0.6$ for all participants, and only 10 (randomly generated) realizations of the AR1 process. Each participant is randomly assigned to one of the 10 paths. Other aspects of the experimental procedures are the same as Experiment 1. There are 330 participants in total, with about 30 participants per path (again the randomization is not perfectly even). Conditions are summarized in Table 1, Panel B.

Experiment 3. In Experiment 3, we modify Experiment 1 in several ways to perform various robustness checks. Table 1, Panel C, provides a summary of the conditions. Every participant is randomly assigned to one of these conditions. There are 875 participants in total, with roughly 35 participants per condition. Experiment 3 was conducted in June 2017.

The conditions in Experiment 3 are designed to help us implement three main tests (whose results we report in Section 5.7):

1. *Well-known economic variables vs abstract process*

   In Conditions 1 to 8, we test whether participants' forecasting behavior is different when we provide a "context" for the random process they forecast. Specifically, we estimate the properties of four major economic variables (assuming an AR1 process): U.S. quarterly GDP growth, monthly CPI, monthly S&P 500 stock returns, and monthly house price growth. We then use the estimated parameters to generate the random processes in the experiment. In Conditions 1 to 4, in the experimental instruction we explain that "the process you will see has the same property as [...]." In Conditions 5 to 8, we use the same random processes but do not provide the "context" in the experimental instruction. Everything else is the same as Experiment 1. Through this design, we can examine whether participants' behavior

is influenced by the "context" by comparing Conditions 1 to 4 with their counterparts in Conditions 5 to 8.

2. *Varying other AR1 parameters than ρ: Long-term mean and volatility*

   In addition, we use Conditions 5 and 6, which both have $\rho = 0.4$ but different values for μ and σ, to test whether these other parameters affect our results. In particular, we compare them to Condition 9, which has $\rho = 0.4$ and $\mu = 0$, $\sigma = 20$ as in Experiment 1.

3. *The term structure of expectations*

   In the remaining conditions of Experiment 3, we test the impact of asking participants to report the term structure of expectations. In conditions 10 to 13, we ask for the $t+1$ forecast only. In conditions 14 to 17, we ask for the $t+2$ forecast only. In conditions 18 to 21, we ask for the $t+1$ and $t+5$ forecasts. Finally, in conditions 22 to 25, we ask for $t+1$ and $t+2$ forecasts, but remove the grey dot that shows the $t+2$ forecast from the previous round, i.e. $F_{t-1}x_{t+1}$.

4.2 Payments

Each participant is offered a base payment of $1.80. In addition to the base payment, participants also receive incentive payments that depend on their performance in the prediction task. Specifically, for each prediction, the participant receives a score that is a decreasing function of the forecasting mistake as for instance in (Dwyer et al., 1993a; Hey, 1994a):

$$S = 100 \times \max(0, 1 - |\Delta|/\sigma) \tag{5}$$

where Δ is the difference between the prediction and the actual realization, and σ is the volatility of the noise term ϵ. For each round, the score is between 0 and 100. We calculate the cumulative score of each participant, and convert it to dollars by dividing 600. The total score is displayed on the top left corner of the prediction screen, and the score associated with each of the past

prediction (if the actual is realized) is displayed at the bottom of the screen (see Figure 1).

For one particular condition (Baseline, AR1 process with $\rho = .6$), we show in Figure 2 the joint distribution of scores and payments. Each point on this figure corresponds to the score (x-axis) and payment (y-axis) of one participant (there are 38 participants in this condition). All points are on straight line as payment is equal to score divided by 600 plus \$1.80. The expected incentive payment for a (full information) rational agent about \$5 (6.25 cents per prediction times 80 predictions). As can be seen from the figure, all agents receive payments that are below the expected rational level. Across all conditions like in the Figure, incentive payments have a mean of approximately \$3.20 (except in those conditions in Experiment 3 where participants give only one forecast per round, where the mean is roughly half as large). Table 2 shows the summary statistics of the incentive payments for all three experiments separately, and within experiment 3, splits between conditions with one and conditions with two forecasts. Clearly, the distribution of payments is on the left of the expected rational bonus of about \$5.

Notice that the loss function defined in equation (5) ensures that a rational participant will choose the rational expectation as an optimal forecast, if there is no cost to do so. This is similar to the earlier experimental literature on expectations Dwyer et al. (1993a); Hey (1994a). $E(1 - |x_{t+1} - F_t|/\sigma)$ is maximal for a forecast F_t equal to the 50^{th} percentile of the distribution of x_{t+1} conditional on x_t. Given that our process is symmetrical around the rational forecast, the median is equal to the mean, and the optimal forecast is therefore equal to the rational expectation.

Finally, the participation constraint of subjects is likely to be satisfied. The sum of the base payment and the incentive payment is about \$5 (for a roughly 15 minute task), which is high compared to the average pay rate on MTurk. As far as the incentive compatibility constraint is concerned, the question is more difficult as it depends on the cost of making more rational expectations. Hozever, recent work by Dellavigna and Pope (2017) show that participants provide high effort even when the size of the incentive payment is modest, and incentives do not appear to be a primary issue in this setting.

4.3 Descriptive Statistics

The distribution of duration is also presented in Table 2. The mean duration of participation is about 13 minutes, and we allow a maximum duration of 60 minutes. The mean duration for each round of prediction is about 10.5 seconds.

Table 3 shows the demographics of participants in our experiments. About 55% to 60% of the participants are male. Roughly 75% report they have college or graduate degrees, and the level of education is higher than that in the general US population (60% with college degrees or above) (Ryan and Bauman, 2015). 40% report they have taken a statistical class. The median age is about 33, slightly lower than the general population (37) (Howden and Meyer, 2011); less than 2% of the participants are above 65.

5 Empirical Results

5.1 Measuring rational expectations

To estimate our econometric specification, we need to compute the rational expectation of the agent, which we generically denote $E_{t-k}x_t$. We use two different measures, which we describe here. The first measure assumes that the agent knows the data-generating process. This corresponds to the full information rational expectation used in most economic models. We thus define rational expectation about x_t conditional on information available at date $t-k$ as:

$$E^{FI}_{t-k}x_t = \rho^k x_t$$

This definition of full information rational expectations will be our baseline, and for simplicity we will use it in most of our regressions.

The participant does not, however, know the data-generating process, so in practice the participant will try to infer it using the data. In robustness checks, we use a definition based on least

square learning Evans and Honkapohja (2001):

$$\widehat{E}_{t-k}x_t = a_{t-k} + \sum_{i=0}^{i=n} b_{i,t-k}x_{t-k-i}$$

whereby, every period, the participant forecasts x_t using all lagged values from x_{t-k-n} until x_{t-k}. Parameters a_{t-k} and $(b_{i,t-k})_i$ are estimated using OLS and all the available past history of realizations of x_t until x_{t-k}. Because of the central limit theorem, a LS learning participant with an infinite number of data points would form full information rational expectations: $a_{t-k} = 0$, $b_{0,t-k} = \rho^k$, $b_{i,t-k} = 0$ for $i > 0$. In the paper, we set $n = 3$ but our results are insensitive to this threshold.

For the $AR1$ processes that we are using in our experiment, the difference between the two definition is not large. In Figure 3, we plot $E^{FI}_{t-1}x_t$ against $\widehat{E}_{t-1}x_t$ for all realizations in our data for which $0 \leq \rho < 1$. As is apparent in the picture, the correlation between the two measures is high: .84. The slope coefficient is .86, so that the two measures are highly correlated and similar in the sample we are looking at. We also show in Appendix Table B.1 that the mean squared difference between these two expectations does not decrease very fast during the time the experiment takes place. This is mostly because the experiment only starts after 40 observations, so the estimated model is already quite precise. As is well known, such similarity would not hold for a more complex data-generating process, e.g. with non-linear terms. Hence, in our experimental setting, there is little scope for learning if participants were rational. There might, however, be scope for learning if participants are not rational LS learners. We return to this issue below.

5.2 Main Result

We now turn to our main result. As discussed in Section 3, we run the following regression on the sample of all participants for which the persistence parameter $\rho \in \{0, .2, .4, .6, .8\}$. For individual i at date t:

$$F_t^i x_{it+1} - E_t x_{it+1} = \lambda \left(F_{t-1}^i x_{it+1} - E_t x_{it+1} \right) + \gamma (x_{it} - E_{t-1} x_{it}) + u_{it+1} \tag{6}$$

where the rational expectation $E_t x_{t+1}$ is in general measured with the full information definition $E_{t-k}^{FI} x_t$, except when noted. We use OLS and cluster the error terms u_{it+1} at the individual i level in order to account for the fact that forecast errors may be autocorrelated at the individual level.

Results are reported in Table 4. Most columns use the full information definition of expectations ($E_{t-1}^{FI} x_{it} = \rho x_{it-1}$) unless otherwise noted. Column 1 assumes no extrapolation ($\gamma = 0$). Expectation appear to be sticky with a coefficient $\lambda = .11$, strongly significant statistically (t stat of 5). An econometrician ignoring potential extrapolation would thus infer that expectations are "11%" sticky and "88%" rational. Note that this coefficient is in the ballpark of estimates of expectation stickiness in the literature (Coibion and Gorodnichenko (2015), Bouchaud et al. (2016) for instance). These estimates use field data and focus on consensus – not individual – forecasts. Column 2 makes the opposite exercise, assuming pure extrapolation, and indeed find evidence of extrapolation, with $\gamma = .36$, significant with a t-stat of 14. The two drivers are included together in column (3), which is our preferred specification. Compared to columns 1 and 2, both γ and λ increase, which is consistent with intuition. They are both very significant. Column 5 confirms our main finding by using the LS learning rational expectation instead of FI expectations. Estimates barely change. All in all, across all specification, γ hovers between .42 and .44. λ hovers between .18 and .21. We will return to magnitudes below.

We also investigate the possibility that learning takes place, reducing systematic errors over time. We do not find much evidence of learning during the 40 periods of our test. One first way of looking at this consists of splitting the sample between the first 20 and the last 20 rounds of testing. If learning takes place, we should see a reduction in the estimated γ and λ. We do this in Table 4, columns 6-7. Stickiness λ decreases slightly, and over-reaction γ increases a bit, but none of these changes are statistically significant. Another way to explore learning in this context consists of computing the mean squared difference between subjective forecasts and

rational forecasts. We show this statistic in Figure 4. In panel A, we show the square root of the mean squared difference between the observed forecast $F_{t-1}x_t$ and the full information rational expectation $E_{t-1}x_t = \rho x_{t-1}$. We plot this number as a function of the round of observation in Panel A. If all participants were rational, this mean difference would be zero. Since we cannot expect participants to be FI rational (they need to learn about the model from the data), we replicate the same analysis with LS expectations, and show the reduction in the distance between LS and FI expectation in Panel B. Clearly, the distance of subjective forecasts to FI expectations is much bigger than LS learners (about 4 times bigger). Also, while this distance is reduced by about 30% after 40 rounds for LS learners, it goes down by less than 10% for the participants of our experiment. The bulk of the downward sloping learning curve is accounted for by the first couple of periods, after which essentially no learning seems to take place.

5.3 Robustness of the estimates of γ and λ

Table 5 offers further evidence that our estimates of γ and λ are very robust across subpopulations. In Panel A, we split the sample of participants by demographic category: Gender (columns 1-2), Age (columns 3-4) and Education (columns 5-6). In Panel B, we split the sample of participants by response to basic questions designed to test the statistical skill of participants. In columns 1-2, we focus on the "coin toss" question, designed to test if participants understand the notion of statistical independence. In columns 3-4, we look at answers to a question designed to see if participants know what a median is. In columns 5-6, we split participants into those who answered right or wrong to the "hospital" questions, which tests if people understand the law of large numbers. In all these subsamples, the stickiness estimate is strongly statistically significant and hovers between .17 and .26. The extrapolation parameter is even more significant and hovers between .41 and .47. Interestingly, our measures of statistical skill have very little effect on the estimates. Among demographics, age is the most discriminating variable, with younger participants significantly less sticky and less extrapolating.

Table 6 reports the estimation of equation (6) for each value of ρ between 0 and 1. For all

stationary processes (i.e. for ρ between 0 and 0.8) the model turns out to be remarkably stable. The stickiness coefficient lies between .12 and .21, but the Fisher test cannot reject the null that all coefficients are equal (p value of .33). The same result arises for the trend parameter, which is estimated across conditions between .38 and .48, but then again, the null that all coefficients are equal across conditions is not rejected (p value of .69). The picture actually remains the same when one includes the condition where $\rho = 1$, but the estimates increase a lot: When the process is actually non stationary, participant expectations become both stickier and more extrapolative. It looks like non-stationary processes are harder to cope with. This change is apparently large, though not significant statistically.

5.4 Quantifying the results

In order to shed some light on the relative importance of extrapolation and stickiness in the dynamics of expectation, we start from the equivalent formulation:

$$F_t x_{t+1} = (1-\lambda)\sum_{k=0}^{\infty} \lambda^k E_{t-k} x_{t+1} + \gamma \sum_{k=0}^{\infty} \lambda^k \left(x_{t-k} - E_{t-k-1} x_{t-k}\right) \tag{7}$$

This formulation is equivalent to our main recursive model (6). We estimate it separately in Appendix A.1 and show the resulting coefficient to be similar to the ones we obtain in Table 4. As is apparent from the above equation, stickiness has itself an effect on extrapolation, through the second term. Current forecasts do not only take into account the recent surprise $x_{t-1} - E_{t-2}x_{t-1}$ but in principle all surprises before this, with exponentially decreasing weights. Hence, there is no extrapolation unless $\gamma > 0$ but large values of λ render extrapolation "sticky" and therefore more effective. Note the similarity of the second term with the expectation model in Barberis et al. (2015). We test this model in Appendix A.1 by regressing $F_t x_{t+1}$ on terms in $E_{t-k}x_{t+1}$ and $x_{t-k-1} - E_{t-k-2}x_{t-k-1}$, and find values consistent with exponentially decreasing parameters.

We report the impulse response of this belief formation process in Figure 5. We show three lines. The first line is the impulse itself x_t: $x_0 = 1$ and $x_t = \rho x_{t-1}$ for $t \geq 1$. The second line is

the rational expectation. The rational agent is first surprised by the impulse: $E_0 x_1 = 0$, then, the rational expectation is given by $E_{t-1} x_t = \rho x_{t-1}$ which is equal to the realization in our impulse response setting. The third line is the simulated forecast using formula (7) for $\gamma = .45$, $\lambda = .2$ and $n = 1$. From Figure 5, it appears quite clearly that our forecasters over-react to the impulse when compared to rational forecasters.

In fact, overreaction clearly dominates given our estimates. Taking into account that x_t follows an AR1 process of persistence ρ, and assuming for conveinience $n = +\infty$ we can rewrite the expected error as:

$$F_t x_{t+1} - E_t x_{t+1} = \sum_{k=0}^{\infty} \underbrace{\lambda^k (\gamma - \lambda \rho^{k+1})}_{a_k} \epsilon_{t-k}$$

for which each term a_k is positive as long as $\lambda < \gamma$, which is consistently the case across our estimates. Hence, our forecasters become over-optimistic as soon as a positive shock hits the process, and remain so forever on average, even though their upward bias goes towards zero. This expression also makes clear that sequences of positive news lead to even more extrapolation, a bit like in Barberis et al. (2017). Given our parameter values. The only situation when agents are underreacting is when, say, a positive signal follows a long sequence of negative ones. In this case, the cumulative extrapolation on past negative shocks dominates the extrapolation on the more recent shock, and, overall, the agent underreact.

5.5 Heterogeneity

In this Section, we explore the heterogeneity that is behind our average model of belief formation. To explore this, we go back to our main model, but allow the coefficients λ and γ to vary across individuals:

$$F_t^i x_{it+1} - E_t x_{it+1} = \lambda_i \left(F_{t-1}^i x_{it+1} - E_t x_{it+1} \right) + \gamma_i (x_{it} - E_{t-1} x_{it}) + u_{it+1} \tag{8}$$

i.e., we run one such regression per subject.

In Figure 6, we show the distributions of stickiness and extrapolation. Two messages emerge. First, the null hypothesis of full information rationality ($\lambda = \gamma = 0$) is rejected at 10% for 215 out of 270 subjects. Second, there is significant dispersion, but a lot of it is due to the small number of observations (39) used to estimate each parameter. To assess the heterogeneity separately from estimation noise, for each individual, we compute the p-value of a test of the null that $\lambda_i = \lambda$, taking into account the fact that both numbers are estimated (we run the two regression using the SURE approach). Individual stickiness differs from average at 5% for only 94 subjects out of 270. Individual extrapolation differs from average for 89 individuals. Overall, we cannot reject about two thirds of the individuals behave like the average model. Expectation formation is quite homogeneous.

We have nonetheless investigated the cross-sectional properties of individual parameters. To do this, we have regressed individual λ_i and γ_i on various characteristics. Consistently with our robustness checks in Tables 5 and 6, we did not find any significant and consistent relation between either of the two parameters and sociodemographics, measures of statistical literacy, or the level of ρ. λ and γ are consistent across these groups. The only cross-sectional relation that emerges if the negative correlation between λ_i and γ_i which is equal to $-.2$ and significant at 5%. Hence, sticky subjects tend to extrapolate less.

5.6 Individual-level vs Consensus Forecast

In this Section, we ask if our model of expectation formation does well to explain aggregate expectations. The model estimated in Table 4 has an R^2 of .15, which suggests that the error term u_{it+1} in equation (6) is quite volatile. This means individual expectations are hard to predict, but average expectations may be easier to predict if some of these errors are idiosyncratic. To check this, we need to make sure that several subjects in a condition are exposed to the same realization.

We start with a slightly different experimental setting. We use a single process with $\rho = .6$. We

then randomly sort subjects into 10 different conditions. Each condition has a different realization of the process, but within each condition, all subjects see the same realization. We then take the average expectations within each condition, and test our model within it. More specifically, we run the following regression:

$$F_t^c x_{ct+1} - E_t x_{ct+1} = \lambda \left(F_{t-1}^c x_{ct+1} - E_t x_{ct+1} \right) + \gamma (x_{ct} - E_{t-1} x_{ct}) + v_{ct+1}$$

for condition c at round t. $F_t^c x_{ct+1}$ is the *average* prediction across subjects in condition c at round t for next period realization. Hence, the panel on which we run these regressions is somewhat smaller than in Table 4. We only observe 40 round in 10 different conditions, hence at most 400 observations in total (vs about 6,000 in our main setting).

We report the results of this regression in Table 7 using the same structure as in Table 4: With each component of the regression separately, with LS rational expectations and full information ones, for the first and last 20 periods separately. Three salient points emerge. First, the coefficients obtained in this setting are very similar to the coefficients obtained in our main specification (.21 vs .19 for stickiness λ and .46 vs .43 for extrapolation γ). Second, the R^2 of this regression (.57) is much higher than in our main specification (.15). Thus, a big part of the error term u_{it} in the individual expectation model are idiosyncratic errors that vanish in aggregation. And overall, our model does a very good job at explaining the expectation formation process. Third, taking LS rational expectation – compared to FI rational expectations – makes a small difference at the aggregate level. The model with LS rational forecast has a higher R^2 (.66 in column 5) than the model with FI expectations. Also, the model with FI expectations works better for the last 20 rounds than in the first 20 rounds of experimentation. Both these are consistent with the idea that LS learning is more realistic.

5.7 Robustness to Experimental Setting

In this last Section, we investigate the robustness of our results to changes in the experimental setting.

5.7.1 Well-known economic variables vs abstract process

First, we check if subjects behave differently when they are forecasting the process of an "actual" economic variable. We focus on four different variables: U.S. quarterly GDP growth, monthly CPI inflation, monthly S&P 500 returns and monthly house price growth. For each of these variables, we first estimate the process as AR1 processes, which we then simulate for each participants (each participant receives a different draw of realized innovation). We then randomly allocate subjects to two conditions: In both conditions, subjects are asked to forecast future realizations, but in one of them, they are told at the beginning of the experimental instructions that "The process you will see has the same property as quarterly US real GDP growth in the last three decades" (for the GDP growth time series). We repeat this procedure for the 4 economic variables, and run our main specification of column 3, Table 4, separately in each condition.

We report the results in Table 8. For each of the four economic variables (GDP, inflation, stock market and housing market returns), we report the estimated equation (6) separately for the two conditions with and without process description. We then test equality of estimated γ and λ across the two conditions, and provide p-values in the bottom two lines. For both parameters, and for all four variables, we cannot reject the null hypothesis that the two models are identical. We deduct from this that knowing subjects' priors about the nature of the variable to predict does not strongly affect their forecasting rule.

5.7.2 Varying other parameters than ρ

Second, we check that the estimated γ and λ do not significantly change when we vary the parameters of the process. We do this by exploiting the conditions described above. In Table

6, we started from our main process, for which $Ex_t = 0$ and $\sigma = 20$, and varied ρ from 0 to 1. We showed that λ and γ did not vary significantly across conditions. In this Section, we use the conditions described in the previous Section where we calibrate the process on existing economic variables (GDP growth, CPI inflation, stock market and housing returns). To make things comparable, we focus on the conditions where subjects were not told anything about these processes. In these different conditions, not only ρ, but also Ex_t and σ vary. This allows us to test if the expectation formation equation changes.

In Table 9, we implement these tests. Columns 2-4 report the regression results for the four economic variables (again, in the conditions where subjects are not told how the process was calibrated). Column 1 shows the baseline condition, for which $Ex_t = 0$, $\sigma = 20$ and $\rho = .4$. The numbers differ slightly from Table 6 because this condition was part of our third batch of experiments, along with all results discussed in Section 5.7. Volatility varies widely, from .23 (inflation) to 20 (baseline condition). The long-term mean goes from 0 (baseline) to .55 (stock returns). The p-value of tests of equality between each of the four conditions and the baseline are in the bottom panel of the Table. We can never reject the null hypothesis that the two models are identical. From this, we deduct that our model is robust to parameter changes.

5.7.3 Reporting the term structure of expectations

Our last robustness check is about the effect of reporting the term structure of expectation. A key dimension of our experimental setting is that we ask subjects to provide us with long-term expectations. This may cause under-reaction via anchoring: Because they are asked to report long-term expectations ($F_t x_{t+2}$), subject may have a propensity to under-react to information available at $t+1$ in order to not modify their long-term expectation too much. We investigate this in several experimental conditions, and report the results in Table 10. Our tests suggest that stickiness is indeed affected the reporting of long-term expectations, in the direction of expectations begin stickier. However, the extrapolation coefficient is surprisingly robust across all conditions.

First, in Table 10 columns 1-2, we ask if our visual presentation of past long-term expectations

affects reporting and expectation formation. In our baseline experimental treatment, we assist the participant's memory by figuring, in round t, her past long-term expectation $F_{t-1}x_{t+1}$ as a grey dot on the graphical interface. Thus, when she makes forecasts $F_t x_{t+1}$ and $F_t x_{t+2}$, she sees – via the gray dot – what she had anticipated for x_{t+1} in the previous round. The gray dot helps remember past forecasts, but also may reinforce anchoring of expectations. So we sort individuals into two conditions, both asking for short- and long-term forecasts but in one condition we do not include the gray dot. We run our main specification (6) and report the results in column 1 (baseline condition) and column 2 (baseline condition without gray dot). Obviously, estimates in column 1 are exactly the same as in Table 9, column 1. We test the equality of coefficients in the bottom panel. The extrapolation coefficient γ is not statistically different across both conditions (.49 without gray dot against .47 in the baseline). The stickiness coefficient λ is however significantly higher with a p value of .03. It is equal to $.24^{***}$ in the baseline condition versus a barely significant $.10^{*}$ in the condition without gray dot. Thus, the presence of the gray dot tends to make the subjects significantly more sticky. Note however that the dominant feature of the data – extrapolation – is unchanged by the absence of the gray dot.

Second, we ask if the mere fact of reporting long-term expectations tends to make short term expectation stickier. We implement these tests in columns 3-5. In column 3, we show the baseline condition. In column 4, we report the results of a conditions where subjects are only required to provide short-term expectations, $F_t x_{t+1}$, and not long-term ones $F_t x_{t+2}$. In column 5, we on the contrary analyze a condition where subjects report short-term expectations and very long-term ones $F_t x_{t+5}$. To compare these three conditions, we cannot run our main specification (6) since it requires both $F_t x_{t+1}$ and lagged $F_t x_{t+2}$, which we don't have the two alternative conditions. Instead, we run the lagged equivalent of (6):

$$F_t^i x_{t+1} = (1-\lambda)\sum_{k=0}^{2}\lambda^k E_{t-k}^i x_{it+1} + \gamma\sum_{k=0}^{2}\lambda^k\left(x_{it-k} - E_{t-k-1}^i x_{it-k}\right) + \eta_{it} \tag{9}$$

which is the same equation as in (7) limited to three lags – coefficients are supposed to be negligible

after 2 lags which is the case in the regressions. Notice that the coefficient on $E_t^i x_{it+1}$ is equal to $1-\lambda$, while the coefficient on current innovation $x_{it} - E_{t-1}^i x_{it}$ is equal to γ. This regression is run columns 3-5, and in the bottom panel we test equality on these two coefficients. The first result is that reporting $F_t x_{t+2}$ does not affect extrapolation γ at all but makes expectation formation significantly less sticky ($1-\lambda$ is $.85^{***}$ instead of $.55^{***}$, with a p value of .05). The second result is more intuitive: Asking for very long term expectations $F_t x_{t+5}$ makes expectations significantly stickier than asking for medium term expectations $F_t x_{t+2}$ ($1-\lambda = .55$ compared to .85 in the baseline). In both alternative conditions, however, the γ coefficient is almost unchanged to .5 (instead of .48). Overall, stickiness is affected by elicitation of long-term expectations, but the quantitatively dominant force, extrapolation, is unchanged.

Third, we also ask if eliciting short-term expectations $F_t x_{t+1}$ affects the reporting of long-term expectation $F_t x_{t+2}$. We do this in columns 6 and 7, where we compare the baseline with a condition where subjects are only asked to report $F_t x_{t+2}$. Like in the previous test, since we only have one expectation and not two, we need to use the lag formulation of our test, except that now we seek to explain $F_t x_{t+2}$ (which is present in both conditions) and not $F_t x_{t+1}$. The extension of equation (7) to this case yields:

$$F_{t-1}^i x_{t+1} = (1-\lambda)\sum_{k=0}^{1} \lambda^k E_{t-1-k}^i x_{it+1} + \gamma \sum_{k=0}^{1} \lambda^k \left(x_{it-k} - E_{t-1-k}^i x_{it-k}\right) + \eta_{it} \tag{10}$$

where the coefficient on $E_{t-1}^i x_{it+1}$ is equal to $1-\lambda$ and the coefficient on $x_{it} - E_{t-1}^i x_{it}$ is equal to γ. We run the regression separately for the two conditions in columns 6 and 7. We find that both coefficients are similar across both settings. Long-term expectations do not seem to be too affected by short-term expectation reporting. The stickiness coefficient is marginally affected, in the direction of long-term expectations being stickier when short-term ones are reported, but the p value is on the high side (p=.13).

6 Conclusion

In this paper, we run a large scale experiment to investigate how people form forecasts of a variable when faced with past realizations of that variable. At both the individual and the aggregate levels, find strong evidence of extrapolative bias and of forecast stickiness. We calibrate a simple model that nests rational expectations, in which both biases can coexist. Extrapolation turns out to be quantitatively the most important bias. Interestingly, we find our parameters to be relatively independent of the process statistical characteristics. Stickiness is stronger when agents are reminded in a more salient manner of their past forecasts. Apart from this, we find that context elements and framing of the experiment do not affect significantly our estimations. We also find that agents do not improve the quality of their forecasts over time.

References

Abarbanell, J. S. and V. L. Bernard (1992): "Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior," *The Journal of Finance*, 47, 1181–1207.

Assenza, T., T. Bao, C. Hommes, and D. Massaro (2014): "Experiments on Expectations in Macroeconomics and Finance," *Research in Experimental Economics*, 17, 11–70.

Ball, R. and P. Brown (1968): "An empirical evaluation of accounting income numbers," *Journal of accounting research*, 159–178.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015): "X-CAPM: An Extrapolative Capital Asset Pricing Model," *Journal of Financial Economics*, 115, 1–24.

——— (2017): "Extrapolation and Bubbles," *Journal of Financial Economics*.

Barberis, N., A. Shleifer, and R. Vishny (1998): "A model of investor sentiment," *Journal of financial economics*, 49, 307–343.

Beshears, J., J. J. Choi, A. Fuster, D. Laibson, and B. C. Madrian (2013): "What goes up must come down? Experimental evidence on intuitive forecasting," *The American economic review*, 103, 570–574.

Bloomfield, R. and J. Hales (2002): "Predicting the next step of a random walk: experimental evidence of regime-shifting beliefs," *Journal of financial Economics*, 65, 397–414.

Bordalo, P., N. Gennaioli, R. Laporta, and A. Shleifer (2017a): "Diagnostic Expectation and Stock Returns," Tech. rep.

Bordalo, P., N. Gennaioli, and A. Shleifer (2017b): "Diagnostic Expectations and Credit Cycles," *forthcoming Journal of finance*.

Bouchaud, J.-P., P. Krüger, A. Landier, and D. Thesmar (2016): "Sticky Expectations and the Profitability Anomaly," Tech. rep.

Casler, K., L. Bickel, and E. Hackett (2013): "Separate but equal? A comparison of

participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing," *Computers in Human Behavior*, 29, 2156–2160.

Cavallo, A., G. Cruces, and R. Perez-Truglia (2016): "Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments," *American Economic Journal: Macroeconomics*, forthcoming.

Cohen, L. and D. Lou (2012): "Complicated firms," *Journal of financial economics*, 104, 383–400.

Coibion, O. and Y. Gorodnichenko (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*.

D'Acunto, F. (2015): "Identity, overconfidence, and investment decisions," Working paper.

De Bondt, W. P. (1993): "Betting on trends: Intuitive forecasts of financial risk and return," *International Journal of forecasting*, 9, 355–371.

Dellavigna, S. and D. Pope (2017): "What Motivates Effort? Evidence and Expert Forecasts," *Review of Economic Studies*.

Dominitz, J. (1998): "Earnings expectations, revisions, and realizations," *Review of Economics and statistics*, 80, 374–388.

Dominitz, J. and C. F. Manski (2011): "Measuring and interpreting expectations of equity returns," *Journal of Applied Econometrics*, 26, 352–370.

Dwyer, G., A. Williams, R. Battalio, and T. Mason (1993a): "Tests of Rational Expectations in a stark Setting," *Economic Journal*.

Dwyer, G. P., A. W. Williams, R. C. Battalio, and T. I. Mason (1993b): "Tests of rational expectations in a stark setting," *The Economic Journal*, 103, 586–601.

Evans, G. and S. Honkapohja (2001): *Learning and Expectations in Macroeconomics*, Princeton University Press.

Frankel, J. A. and K. A. Froot (1985): "Using survey data to test some standard propositions regarding exchange rate expectations," .

Fuster, A., D. Laibson, and B. Mendel (2010): "Natural expectations and macroeconomic fluctuations," *The Journal of Economic Perspectives*, 24, 67–84.

Gabaix, X. and D. Laibson (2001): "The 6D bias and the equity-premium puzzle," *NBER macroeconomics annual*, 16, 257–312.

Gennaioli, N., Y. Ma, and A. Shleifer (2015): "Expectations and Investment," .

Gourinchas, P.-O. and A. Tornell (2004): "Exchange rate puzzles and distorted beliefs," *Journal of International Economics*, 64, 303–333.

Greenwood, R. and A. Shleifer (2014): "Expectations of returns and expected returns," *The Review of Financial Studies*, 27, 714–746.

Hey, J. (1994a): "Expectations formation: Rational or adaptive or ... ?" *Journal of Economic Behavior and Organizations*.

Hey, J. D. (1994b): "Expectations formation: Rational or adaptive or??" *Journal of Economic Behavior & Organization*, 25, 329–349.

Hommes, C. (2011): "The heterogeneous expectations hypothesis: Some evidence from the lab," *Journal of Economic dynamics and control*, 35, 1–24.

Hong, H., T. Lim, and J. Stein (2000): "Bad News Travel Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies," *Journal of Finance*, 55, 265–295.

Hou, K. (2007): "Industry Information Diffusion and the Lead-Lag Effect in Stock Returns," *Review of Financial Studies*, 20, 1113–1138.

Howden, L. M. and J. A. Meyer (2011): "Age and Sex Composition: 2010," Us census bureau report, US Census Bureau.

Ikenberry, D., J. Lakonishok, and T. Vermaelen (1995): "Market underreaction to open market share repurchases," *Journal of financial economics*, 39, 181–208.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015): "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments," *American Economic Review*, 105, 1478–1508.

Lakonishok, J. and I. Lee (2001): "Are insider trades informative?" *The Review of Financial Studies*, 14, 79–111.

Lakonishok, J., A. Shleifer, and R. Vishny (1994): "Contrarian Investment, Extrapolation and Risk," *Journal of Finance*, 49, 1541–1578.

Laporta, R. (1996): "Expectations and the Cross-Section of Stock Returns," *Journal of Finance*, 51, 1715–1742.

Lian, C., Y. Ma, and C. Wang (2017): "Low Interest Rate and Risk Taking: Evidence from Individual Investment Decisions," Working paper.

Mankiw, N. G. and R. Reis (2001): "Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve," Tech. rep., National Bureau of Economic Research.

Nerlove, M. (1958): "Adaptive Expectations and Cobweb Phenomena," *Quarterly Journal of Economics*.

Ryan, C. L. and K. Bauman (2015): "Educational Attainment in the United States: 2015," Us census bureau report, US Census Bureau.

Schmalensee, R. (1976): "An experimental study of expectation formation," *Econometrica: journal of the Econometric Society*, 17–41.

Shiller, R. (1981): "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review*.

Figures

Figure 1: Prediction Screen

Below is a screen shot of the prediction task. The green dots indicate past realizations of the statistical process. In each round t, participants are asked to make predictions about two future realizations $F_t x_{t+1}$ and $F_t x_{t+2}$. They can drag the mouse to indicate $F_t x_{t+1}$ in the purple bar and indicate $F_t x_{t+2}$ in the red bar. Their predictions are shown as yellow dots. The grey dot is the prediction of x_{t+1} from the previous round $F_{t-1} x_{t+1}$; participants can see it but cannot change it. After they have made their predictions, participants click "Make Predictions" and move on to the next round. The total score is displayed on the top left corner, and the score associated with each of the past prediction (if the actual is realized) is displayed at the bottom.

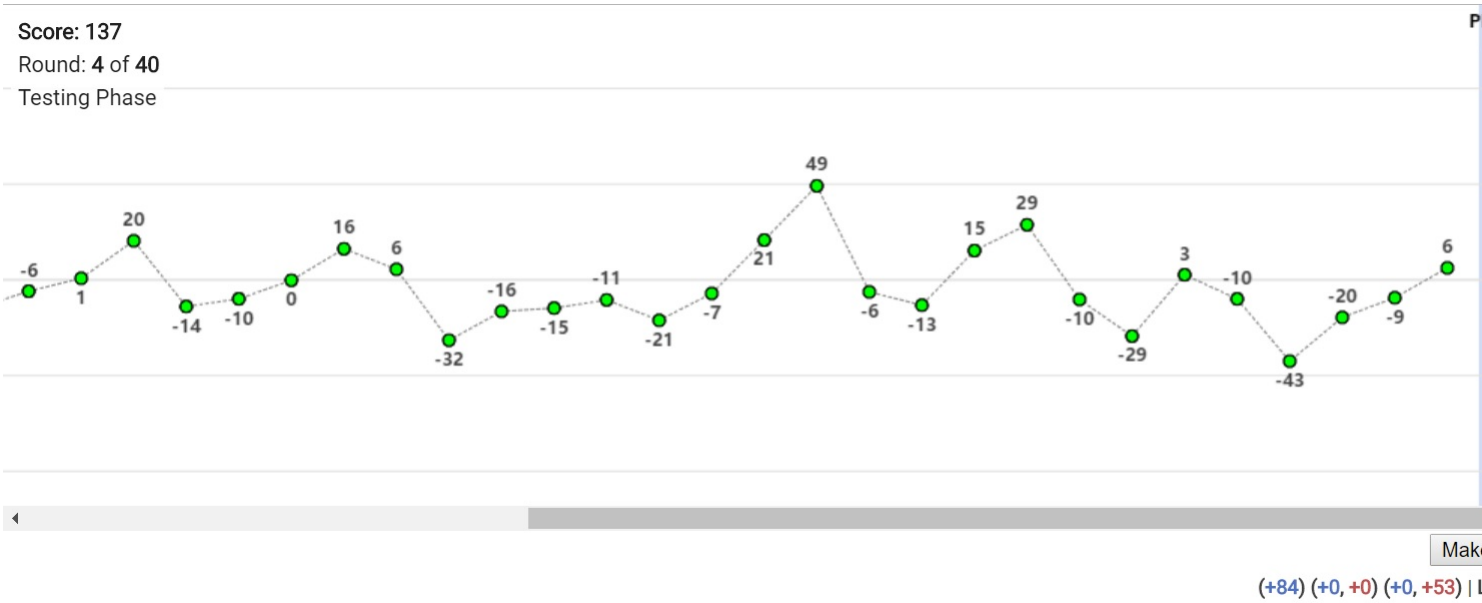
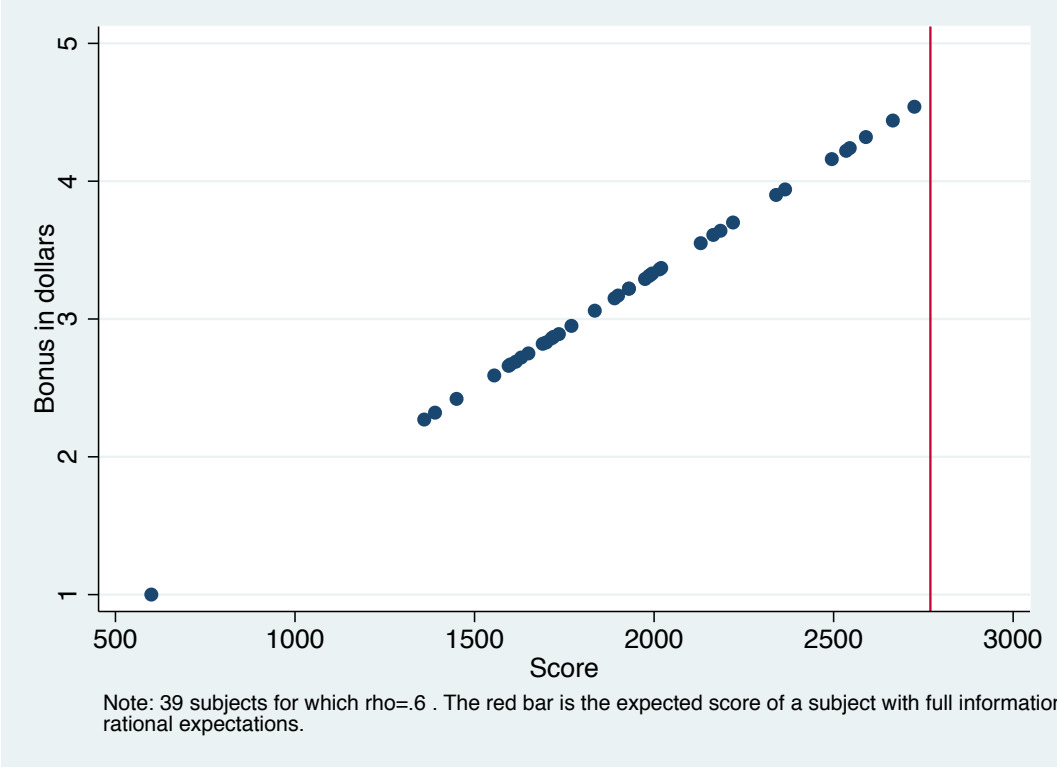
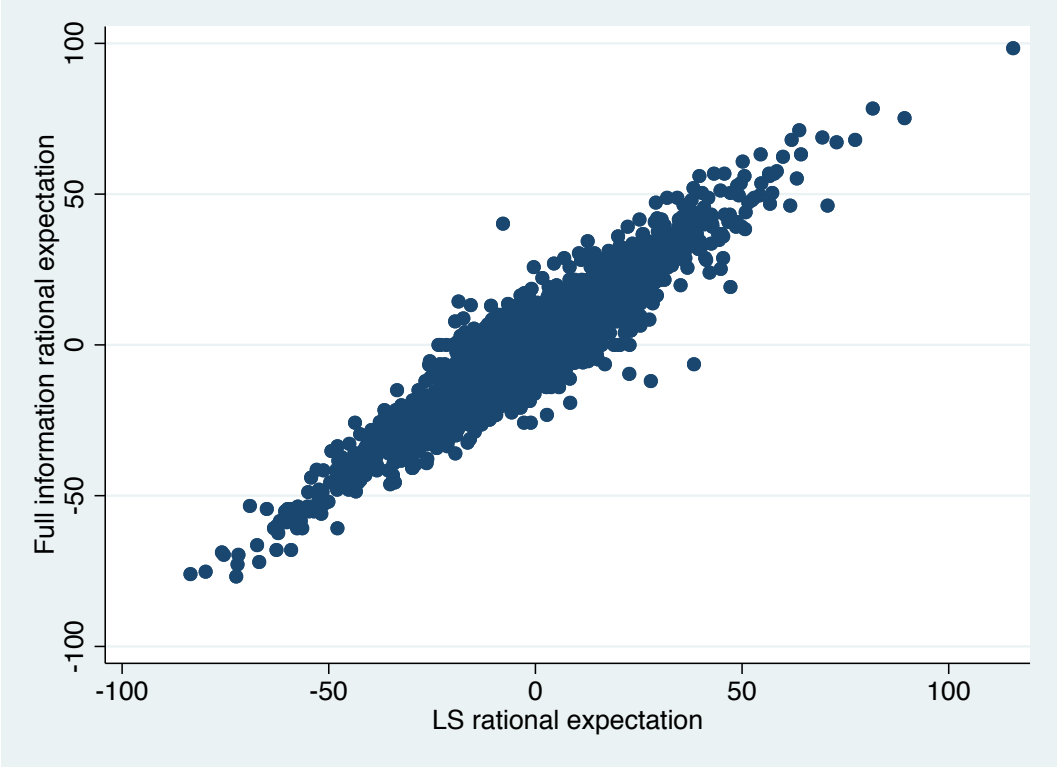


Figure 2: Payment and scores in the Baseline Experiment



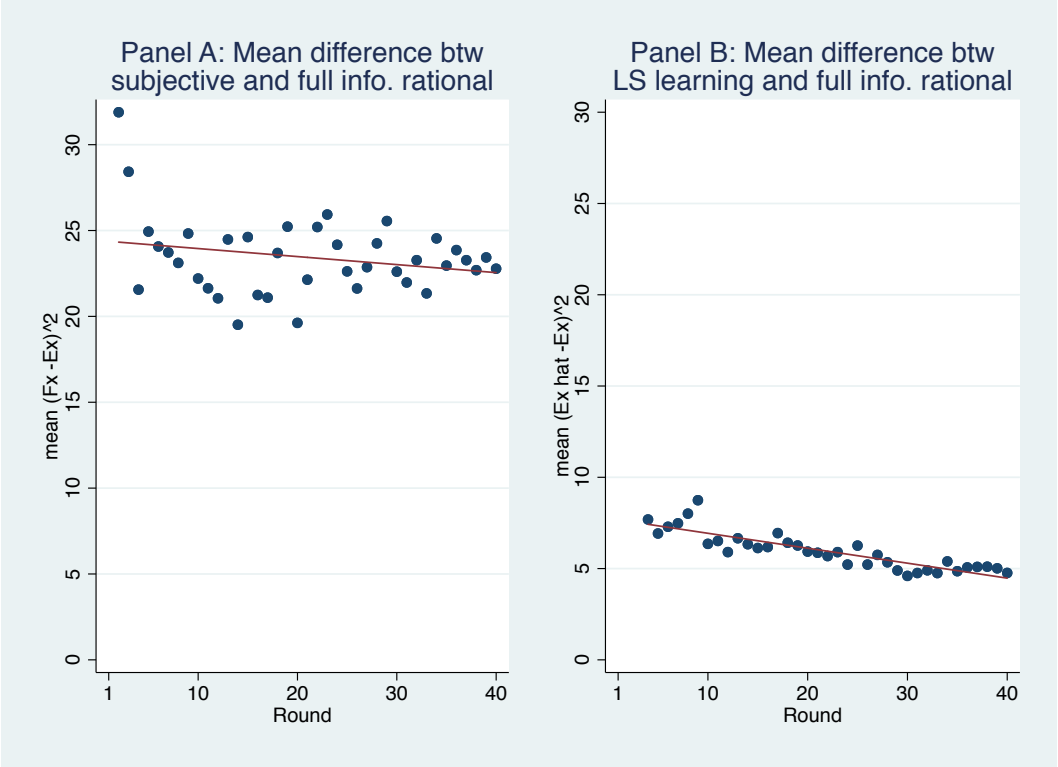
Note: Each point on this figure corresponds to one participant in one condition of the baseline experiment (Experiment 1, with $\rho = .6$). On the x-axis, we report the score obtained, and on the y-axis, the payment in \$, which is equal to the score divided by 600. The vertical red line on the right represents the expected payment of a (full information) rational participant for which $F_t x_{t+1} = E_t x_{t+1} = \rho x_t$.

Figure 3: Full Information vs Least Square Expectations



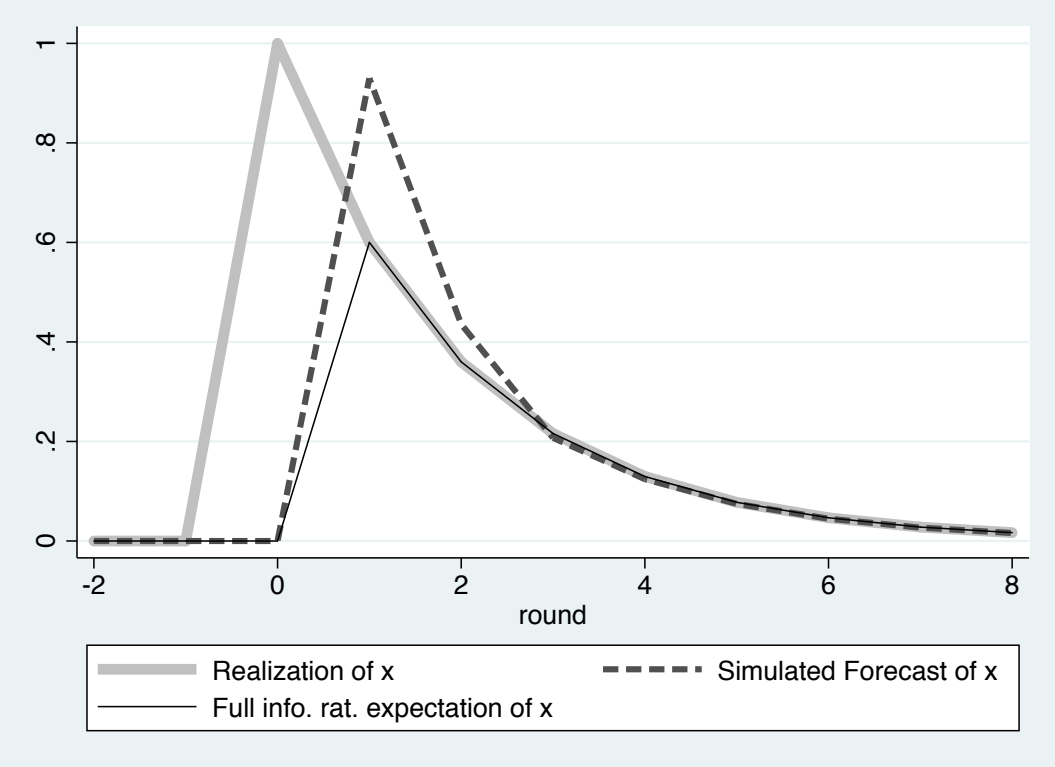
Note: Each point on this figure corresponds to one participant in one testing round. On the x-axis, we report the LS expectations of x_t using three lags $x_{t-1}, x_{t-2}, x_{t-3}$ and coefficients estimated using OLS and all information available until date $t-1$. On the y-axis, we report the FI expectation given by ρx_{t-1}. We only focus on participants for which $\rho \geq 0$ and $\rho < 1$. Regressing y on x leads to an $R^2 = .84$ and a slope coefficient of .86.

Figure 4: How Fast Do Subjective Forecasts Converge to Rational Expectations?



Note: We investigate here the speed at which participants' subjective forecasts converge to full information rational forecasts (Panel A). We compare this to what rational least-square learners would do (Panel B). We use all conditions of the experiment # 1, i.e. all participants with $\rho \in \{0, .2, .4, .6, .8\}$. For each testing round t from 1 to 40, we compute the mean square difference between the subjective forecast $F_{t-1}x_t$ and the full information rational forecast $E_{t-1}x_t = \rho x_{t-1}$. We then take the square root of this, and report it in Panel A. Hence, in Panel A, if all survey participants were full information rational, the mean difference would be equal to zero. We then repeat this procedure in Panel B, replacing the subjective forecast with the LS learning expectation $E^{LS}_{t-1}x_t$ obtained by regressing x_s on x_{s-1} for all periods between -40 and $t-1$. Hence, Panel B allows to observe the extent to which a LS learner would converge to the true rational expectation. Reading: The root mean squared difference between FIR and subjective forecasts goes down from 24 to 22 after 40 rounds. The root mean squared difference between FIR and LS learner forecasts goes down from 7 to 5 after 40 rounds.

Figure 5: Expectation Response to an Impulse in x



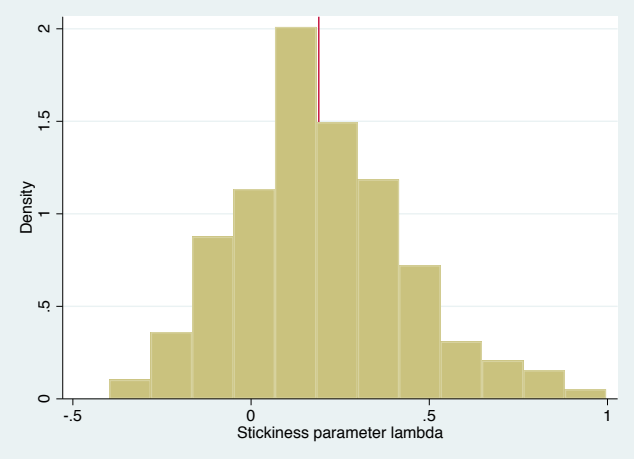
Note: We assume $\rho = .6$ and show the impulse response of a process x, its rational expectation and its forecast using the formulation estimated in this paper. The thick light grey line correspond to the simulation of the response of an AR1 x_t to a one time shock in ϵ equal to 1. Hence, $x_0 = 1$ and for each $t \geq 1$, $x_t = .6x_{t-1}$. The fine dark line is the full information rational expectation, equal to $E_{t-1}x_t = 0$ until $t = 0$, and equal to $E_{t-1}x_t = \rho x_{t-1}$ for $t \geq 1$. The dark dashed line corresponds to the forecasting process estimated in the paper. We use the lag formulation described in Section 5.4 and estimated in Appendix A.1:

$$\begin{aligned} F_{t-1}x_t \approx\ & .8E_{t-1}x_t + .16E_{t-2}x_t \\ & + .45\left(x_{t-1} - E_{t-2}x_{t-1}\right) + .09\left(x_{t-2} - E_{t-3}x_{t-2}\right) \end{aligned}$$

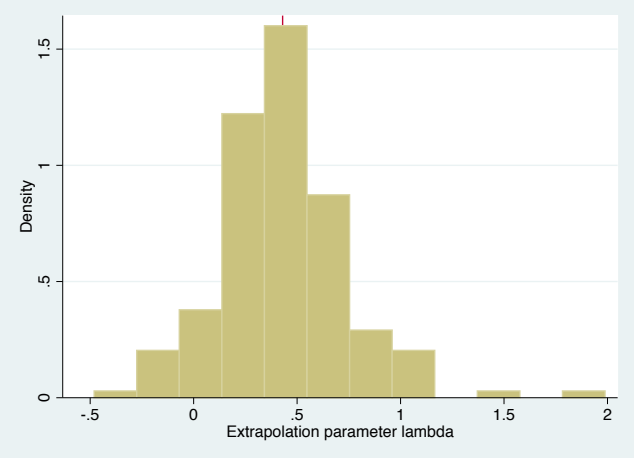
for all $t \geq 1$.

Figure 6: Sample distribution in Stickiness and Extrapolation

Panel A: Distribution of Stickiness λ



Panel B: Distribution of Stickiness γ



Note: On the panel of participants for which $\rho \in \{0, .2, .4, .6, .8\}$, we run the following regression:

$$F_t^i x_{it+1} - E_t x_{it+1} = \lambda_i \left(F_{t-1}^i x_{it+1} - E_t x_{it+1} \right) + \gamma_i (x_{it} - E_{t-1} x_{it}) + u_{it+1}$$

where λ_i and γ_i are allowed to differ across subjects. We then report the distribution of these parameters in the two panels above. The vertical red line corresponds to the estimates of the average model in Table 4, column 3.

Tables

Table 1: Summary of Conditions in all Three Experiments

#	Short description	(1) AR1 process persistence ρ	(2) AR1 process Lt mean μ	(3) Volatility σ_ϵ	(4) Forecasts asked	(5) Grey dot	(6) Number of participants
Panel A: Experiment 1 – Baseline							
A1	Baseline $\rho = 0$	0	0	20	F1+F2	Y	32
A2	Baseline $\rho = 0.2$	0.2	0	20	F1+F2	Y	32
A3	Baseline $\rho = 0.4$	0.4	0	20	F1+F2	Y	36
A4	Baseline $\rho = 0.6$	0.6	0	20	F1+F2	Y	39
A5	Baseline $\rho = 0.8$	0.8	0	20	F1+F2	Y	28
A6	Baseline $\rho = 1$	1	0	20	F1+F2	Y	40
Panel B: Experiment 2 – Common path							
B1	Path 1	0.6	0	20	F1+F2	Y	37
B2	Path 2	0.6	0	20	F1+F2	Y	32
B3	Path 3	0.6	0	20	F1+F2	Y	37
B4	Path 4	0.6	0	20	F1+F2	Y	30
B5	Path 5	0.6	0	20	F1+F2	Y	32
B6	Path 6	0.6	0	20	F1+F2	Y	33
B7	Path 7	0.6	0	20	F1+F2	Y	27
B8	Path 8	0.6	0	20	F1+F2	Y	33
B9	Path 9	0.6	0	20	F1+F2	Y	26
B10	Path 10	0.6	0	20	F1+F2	Y	43
Panel C: Experiment 3 – Robustness checks							
C1	context: quarterly GDP growth	0.4	0.40	0.55	F1+F2	Y	38
C2	context: monthly inflation	0.4	0.12	0.23	F1+F2	Y	39
C3	context: monthly stock returns	0.2	0.55	3.43	F1+F2	Y	29
C4	context: monthly house price growth	0.8	0.02	0.39	F1+F2	Y	37
C5	no context, comparison	0.4	0.40	0.55	F1+F2	Y	30
C6	no context, comparison	0.4	0.12	0.23	F1+F2	Y	34
C7	no context, comparison	0.2	0.55	3.43	F1+F2	Y	36
C8	no context, comparison	0.8	0.02	0.39	F1+F2	Y	35
C9	comparison	0.4	0	20	F1+F2	Y	30
C10	change horizon	0.2	0	20	F1	/	37
C11	change horizon	0.4	0	20	F1	/	36
C12	change horizon	0.6	0	20	F1	/	33
C13	change horizon	0.8	0	20	F1	/	38
C14	change horizon	0.2	0	20	F2	Y	38
C15	change horizon	0.4	0	20	F2	Y	51
C16	change horizon	0.6	0	20	F2	Y	32
C17	change horizon	0.8	0	20	F2	Y	42
C18	change horizon	0.2	0	20	F1+F5	Y	27
C19	change horizon	0.4	0	20	F1+F5	Y	34
C20	change horizon	0.6	0	20	F1+F5	Y	29
C21	change horizon	0.8	0	20	F1+F5	Y	41
C22	no grey dot	0.2	0	20	F1+F2	N	26
C23	no grey dot	0.4	0	20	F1+F2	N	31
C24	no grey dot	0.6	0	20	F1+F2	N	30
C25	no grey dot	0.8	0	20	F1+F2	N	42

Note: This Table provides a synthetic description of the three experiments we conducted. Each panel is devoted to one experiment, and within each panel, each line corresponds to one experimental condition The first three column (1)-(3) gives the parametrization of the AR1 process $x_{t+1} = \rho x_t + (1-\rho)\mu + \epsilon_{t+1}$ participants are asked to forecast. Column (4) shows the forecasts asked to each participants. "F1+F2" means one- and two-period ahead forecasts. Column (5) indicates if a grey dot is present on the interface to help the participant memorize the long-term forecast of the previous period. Column (6) reports the number of participants. Typically, each participant is presented with a different draw, except in experiment B, where all participants within a given condition are presented with the same draw. Participants were not allowed to participate to several experiments.

Table 2: Experimental Statistics

	Mean	p25	p50	p75	SD	N
Experiment 1 (2 forecastsper round)						
Total time (min)	13.88	8.30	11.65	16.42	8.65	270
Forecast time (min)	7.10	4.49	5.77	7.85	4.39	270
per round (sec)	10.64	6.74	8.66	11.77	6.59	270
Bonus ($)	3.33	2.80	3.31	3.84	0.78	270
Experiment 2 (2 forecasts per round)						
Total time (min)	13.12	8.16	10.88	15.79	7.88	330
Forecast time (min)	6.78	4.59	5.75	7.74	4.05	330
per round (sec)	10.17	6.89	8.63	11.61	6.07	330
Bonus ($)	3.22	2.74	3.15	3.67	0.84	330
Experiment 3 (2 forecasts per round)						
Total time (min)	12.44	7.83	10.56	14.60	7.48	580
Forecast time (min)	6.69	4.42	5.73	7.67	4.17	580
per round (sec)	10.03	6.63	8.59	11.50	6.26	580
Bonus ($)	3.29	2.78	3.29	3.80	0.81	580
Experiment 3 (1 forecast per round)						
Total time (min)	11.05	6.91	9.25	13.73	6.52	295
Forecast time (min)	5.27	3.36	4.31	6.03	3.80	295
per round (sec)	7.91	5.03	6.47	9.05	5.70	295
Bonus ($)	1.62	1.31	1.62	1.92	0.48	295

Table 3: Sample Demographics

		N	%	N	%	N	%
Gender	Male	151	55.9	201	60.9	457	52.2
	Female	119	44.1	129	39.1	418	47.8
Age	<= 25	36	13.3	44	13.3	129	14.7
	25-45	186	68.9	224	67.9	593	67.8
	45-65	44	16.3	57	17.3	145	16.6
	65+	4	1.5	5	1.5	8	0.9
Education	Grad school	26	9.6	42	12.7	121	13.8
	College	170	63.0	200	60.6	524	59.9
	High school	74	27.4	88	26.7	224	25.6
	Below/other	0	0.0	0	0.0	6	0.7
Invest. Exper.	Extensive	7	2.6	6	1.8	23	2.6
	Some	71	26.3	74	22.4	193	22.1
	Limited	100	37.0	129	39.1	367	41.9
	None	92	34.1	121	36.7	292	33.4
Taken Stat Class	Yes	110	40.7	144	43.6	406	46.4
	No	160	59.3	186	56.4	469	53.6
Total		270	100.0	330	100.0	875	100.0

Table 4: Main Expectation Formation Model: Main results

	$F_t x_{t+1} - E_t x_{t+1}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sticky	Trend	Main		$\widehat{E_t} x_{t+1}$	$t \leq 20$	$t > 20$
$F_{t-1}x_{t+1} - E_t x_{t+1}$	.1***		.19***		.19***	.2***	.17***
	(4.9)		(8.4)		(8.9)	(6)	(5.9)
$x_t - E_{t-1}x_t$		.37***	.44***	.44***	.45***	.42***	.45***
		(15)	(19)	(16)	(20)	(13)	(14)
$F_{t-1}x_{t+1}$				.19***			
				(8.3)			
$E_t x_{t+1}$				-.21***			
				(-5.3)			
N	6346	6513	6346	6346	6012	3006	3340
r2	.018	.1	.16	.16	.18	.15	.16

Note: On the panel of participants for which $\rho \in \{0, .2, .4, .6, .8\}$, we run the following regression:

$$F_t^i x_{it+1} - E_t x_{it+1} = \lambda \left(F_{t-1}^i x_{it+1} - E_t x_{it+1} \right) + \gamma (x_{it} - E_{t-1} x_{it}) + u_{it+1}$$

In all columns but column (5), we use the FI expectation to measure rational expectations. In column (1), we set $\gamma = 0$. In column (2), we set $\lambda = 0$. Column (3) is our main specification. Column (4) allows γ to differ for both components of the trend regressor. Column (5) uses LS learning rational expectation instead of FI. Columns (6) and (7) split the sample into the first and last 20 rounds. t-stats between brackets.

Table 5: Main Expectation Formation Model:
Sample Splits by Participant Groups

Dependent variable	$F_t x_{t+1} - E_t x_{t+1}$					
Panel A : Socio-demographics						
	(1)	(2)	(3)	(4)	(5)	(6)
	Male	Female	age<35	age≥ 35	high school	college
$F_{t-1}x_{t+1} - E_t x_{t+1}$	.19***	.19***	.15***	.26***	.26***	.17***
	(5.8)	(6.7)	(5.7)	(7.7)	(5.7)	(6.8)
$x_t - E_{t-1}x_t$	.44***	.43***	.42***	.48***	.46***	.43***
	(14)	(13)	(13)	(16)	(11)	(16)
N	3458	2888	3762	2584	1710	4636
r2	.15	.16	.12	.24	.21	.14
Panel B : Answers to statistics quiz						
	(1)	(2)	(3)	(4)	(5)	(6)
	Coin toss		Median		Hospital	
	False	Right	False	Right	False	Right
$F_{t-1}x_{t+1} - E_t x_{t+1}$	.17***	.2***	.18***	.2***	.2***	.16***
	(6)	(6.2)	(5.4)	(6.4)	(9.2)	(3)
$x_t - E_{t-1}x_t$	.44***	.43***	.44***	.44***	.45***	.42***
	(10)	(16)	(13)	(13)	(17)	(9.7)
N	2356	3990	3078	3268	4294	2052
r2	.16	.15	.14	.17	.17	.13

Note: This Table estimates our core regression on subsamples of our experiment. On the panel of participants for which $\rho \in \{0, .2, .4, .6, .8\}$, we run the following regression:

$$F_t^i x_{it+1} - E_t x_{it+1} = \lambda \left(F_{t-1}^i x_{it+1} - E_t^i x_{it+1} \right) + \gamma (x_{it} - E_{t-1}^i x_{it}) + u_{it+1}$$

In Panel A, we focus on sociodemographic categories. In columns 1 and 2, we split the sample into male and female participants. In columns 3 and 4, we split the sample into participants above and below 35 years old. In columns 5 and 6, we split the sample into t-stats between brackets. In Panel B, we focus on groups by answers to various statistical questions. Columns 1-2 split participants into wrong and false answers to the "coin toss" question, designed to see if people understand the notion of statistical independence. Columns 3 and 4 split participants into wrong and false answers to the "median" question, designed to see if people know how to compute a median. Columns 5 and 6 split participants into wrong and false answers to the "hospital" question, designed to see if people understand the law of large number.

Table 6: Main Expectation Formation Model
Sample split by value of ρ

	$F_t x_{t+1} - E_t x_{t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\rho =$	0	.2	.4	.6	.8	1
$F_{t-1}x_{t+1} - E_t x_{t+1}$	.16**	.17***	.12**	.23***	.21***	.43***
	(2.2)	(4)	(2.3)	(6.5)	(6.6)	(3.7)
$x_t - E_{t-1}x_t$	.44***	.48***	.46***	.43***	.38***	.58***
	(8.2)	(7.9)	(11)	(9.3)	(8.5)	(4.8)
N	1216	1216	1368	1482	1064	1520
r2	.17	.2	.14	.15	.13	.29

Note: On the panel of participants for which $\rho \in \{0, .2, .4, .6, .8, 1\}$, we run the following regression:

$$F_t^i x_{it+1} - E_t x_{it+1} = \lambda \left(F_{t-1}^i x_{it+1} - E_t^i x_{it+1} \right) + \gamma (x_{it} - E_{t-1}^i x_{it}) + u_{it+1}$$

In each column, we estimate the above equation for all participants with a given value of ρ. t-stats between brackets.

Table 7: Explaining Average Expectations

	$F_t x_{t+1} - E_t x_{t+1}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sticky	Trend	Main		$\widehat{E_t} x_{t+1}$	$t \leq 20$	$t > 20$
$F_{t-1} x_{t+1} - E_t x_{t+1}$	-.2***		.21***		.3***	.26***	.15***
	(-7.1)		(7.1)		(11)	(6.1)	(3.9)
$x_t - E_{t-1} x_t$		.31***	.46***	.4***	.5***	.45***	.46***
		(12)	(15)	(4.2)	(20)	(9)	(16)
$F_{t-1} x_{t+1}$				.17***			
				(3)			
$E_t x_{t+1}$				-.11			
				(-.8)			
N	380	390	380	380	259	180	200
r2	.15	.48	.57	.57	.66	.48	.68

Note: This Table follows the structure of Table 4, except that the panel data now consists of 10 conditions followed over 40 rounds. In each condition, on average 30 participants make forecast but are exposed to the same draw of a unique AR1 process with persistence $\rho = .6$. Thus, we average forecasts and expectations across participants of each condition. We then run the following regression:

$$F_t^c x_{ct+1} - E_t x_{ct+1} = \lambda \left(F_{t-1}^c x_{ct+1} - E_t x_{ct+1} \right) + \gamma (x_{ct} - E_{t-1} x_{ct}) + u_{ct+1}$$

for condition c at round t. $F_t^c x_{ct+1}$ is the *average* prediction across subjects in condition c at round t for next period realization. The various columns are the same as in Table 4.

Table 8: Sensitivity to Context

	$F_t x_{t+1} - x_t$							
	"GDP growth"		"Inflation"		"Stock returns"		"House price"	
	Without	With	Without	With	Without	With	Without	With
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$F_{t-1}x_{t+1} - E_t x_{t+1}$	.21***	.26***	.3***	.31***	.34***	.19**	.39**	.27***
	(4)	(7.2)	(4.6)	(5.8)	(4.4)	(2.5)	(2.7)	(3.3)
$x_t - E_{t-1}x_t$	.38***	.44***	.37***	.38***	.51***	.45***	.51***	.42***
	(5.9)	(7.8)	(9)	(7)	(8.3)	(6.8)	(4.6)	(4.4)
N	1140	1444	1292	1482	1368	1102	1330	1406
r2	.15	.16	.16	.2	.22	.19	.22	.14
Test of equality – p value								
Stickiness	0.37		0.92		0.18		0.45	
Extrapolation	0.46		0.91		0.51		0.55	

Note: We test here whether "labelling" the process affects the forecasts. For each condition, we run the following regression:

$$F_t^i x_{it+1} - E_t x_{it+1} = \lambda \left(F_{t-1}^i x_{it+1} - E_t^i x_{it+1} \right) + \gamma (x_{it} - E_{t-1}^i x_{it}) + u_{it+1}$$

Columns 1-2 investigate the impact of labelling the process "GDP growth". We estimate delta log GDP as an AR1 on quarterly US data, which leads to $x_t = .40 + .4x_{t-1} + .55\epsilon_t$. We then simulate one path per individual. In column 1, individuals are not told how the process was estimated. This condition is essentially similar – up to a change in innovation volatility and average – to our main tests in Table 6. In column 2, we write at the beginning of the consent form that the process shown replicates that of US GDP growth. We then report p-values of equality tests of λ and γ across samples in the bottom panel of the Table. In columns 3-4, we simulate a process estimated on monthly US CPI inflation. In columns 5-6, we simulate a process estimated on monthly S&P 500 returns. In columns 7-8, we simulate a process estimated on monthly house price growth. Each subject has a different draw of the process, so we cluster error terms at the subject level – thus allowing for within subject correlation of errors but not across subjects. t-stats are between brackets.

Table 9: Sensitivity to Process Parameters Beyond Changes in ρ

Dependent variable			$F_t x_{t+1} - x_t$		
Condition	Main setting	GDP growth	CPI Inflation	S&P returns	House Inflation
Long-term mean $\mu =$	0	0.40	0.12	0.55	0.02
Persistence $\rho =$	0.4	0.4	0.4	0.2	0.8
Innovation vol. $\sigma =$	20	0.55	0.23	3.43	.39
	(1)	(2)	(3)	(4)	(5)
$F_{t-1}x_{t+1} - E_t x_{t+1}$	.24***	.21***	.3***	.34***	.39**
	(4.9)	(4)	(4.6)	(4.4)	(2.7)
$x_t - E_{t-1}x_t$	.47***	.38***	.37***	.51***	.51***
	(9.6)	(5.9)	(9)	(8.3)	(4.6)
N	1596	1140	1292	1368	1330
r2	.23	.15	.16	.22	.22
Test of equality with main setting – p value					
Stickiness	.	0.60	0.51	0.30	0.33
Extrapolation	.	0.23	0.11	0.68	0.77

Note: We test here whether changes in σ, the volatility of innovation, and μ, the long term mean of the process, affect our estimates of λ and γ. In each of these conditions, the process for participant i is given by $x_{it+1} = \rho x_{it} + (1-\rho)\mu + \sigma\epsilon_{it}$ where ϵ_{it} is standardized normal. Each column corresponds to a condition where σ and μ are different. For each condition, we then run the regression:

$$F_t^i x_{it+1} - E_t x_{it+1} = \lambda \left(F_{t-1}^i x_{it+1} - E_t^i x_{it+1} \right) + \gamma (x_{it} - E_{t-1}^i x_{it}) + u_{it+1}$$

Columns 1 is the benchmark model (it slightly differs from Table 4, column 3 because the data come from a different draw of innovations). Column 2 uses the process fitted on US quarterly GDP growth. Column 2 uses parameter from US monthly CPI inflation, column 3 from stock market returns and column 4 from housing returns. These data – and therefore the coefficients – are the same as in Table 8, columns 1,3,5,7. We then report, in the bottom panel, p-value of tests of equality between coefficients of these processes and the benchmark setting of column 1. These tests are done by running the two regressions as SURE. Each subject has a different draw of the process, so we cluster error terms at the subject level – thus allowing for within subject correlation of errors but not across subjects. t-stats are between brackets.

Table 10: Sensitivity to Term Structure Reporting

Dependent variable	$F_t^i x_{it+1} - E_t x_{it+1}$		$F_t x_{t+1}$			$F_{t-1} x_{t+1}$	
	Grey dot		Effect of reporting LT expec.			Effect of reporting ST expec.	
	Main setting	Without	Main setting	$F_t x_{t+1}$ only	$F_t x_{t+1}$ & $F_t x_{t+5}$	Main setting	$F_t x_{t+2}$ only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$F_{t-1} x_{t+1} - E_t x_{t+1}$	.24***	.096*					
	(4.9)	(1.9)					
$x_t - E_{t-1} x_t$	.47***	.49***					
	(9.6)	(7.9)					
$E_t x_{t+1}$			.85***	.51***	.55***		
			(7.5)	(4)	(3.7)		
$E_{t-1} x_{t+1}$			.21*	.36***	.24*	.97***	.63***
			(1.7)	(2.9)	(1.8)	(5.9)	(4.1)
$E_{t-2} x_{t+1}$			.047	.38**	.17**	.22	.63***
			(.56)	(2.5)	(2)	(1.3)	(4.3)
$x_t - E_{t-1} x_t$			.48***	.5***	.5***		
			(8.6)	(7.7)	(6.3)		
$x_{t-1} - E_{t-2} x_{t-1}$			.0058	.021	.05	.45***	.54***
			(.18)	(.7)	(1.6)	(10)	(9.8)
$x_{t-2} - E_{t-3} x_{t-2}$			-.027	-.058***	-.021	.033	.015
			(-1.5)	(-2.9)	(-1)	(1.3)	(.53)
N	1596	1026	4884	5032	4736	4884	5883
r2	.23	.15	.49	.45	.39	.34	.29
Test of equality with main setting – p value							
Stickiness	.	0.03	.	0.05	0.10	.	0.13
Extrapolation	.	0.83	.	0.79	0.80	.	0.18

Note: next page

Sensitivity to Term Structure Reporting (Cont'd)

Note: We test here whether the reporting of both short-term and long-term expectations. In columns 1 and 2, we test whether the presence of a gray dot, to help subjects remember their past two-periods ahead forecast, affects expectation formation. In column 1, we report the results of our baseline setting, identical to Table 9, column 1. In column 2, we use exactly the same parameters and setting but remove the grey dot designed to help subject remembering $F_{t-1}x_{t+1}$ when they need to report F_tx_{t+1} and F_tx_{t+2}. We run our main specification (6) on both conditions and test the equality of coefficients in the bottom panel. In columns 3-5, we test whether the reporting of long-term expectations affects the reporting of short-term ones. In these columns, we use the specification with lags (7), where we regress F_tx_{t+1} on lagged values of rational expectations $E_{t-k}x_{t+1}$ and past innovations of $x_{t-k} - E_{t-k-1}x_{t-k}$ for $k \geq 0$. Under our recursive model (6), the coefficient on the first lags E_tx_{t+1} and $x_t - E_{t-1}x_t$ are equal to $1-\lambda$ and γ respectively. In column 1, we report the baseline condition of column 1, but using the "lag" specification. In column 2, we report the condition where subjects are confronted with the same process, but are not required to report the "long-term" expectation F_tx_{t+2}. In column 3 on the contrary, subjects are required to report very long-term expectations F_tx_{t+5}. We test equality of these coefficients with estimates of column 3 in the bottom panel. In columns 6-7, we test the effect of reporting short-term expectations on long-term ones. The methodology is identical to columns 3-5, except that now we regress $F_{t-1}x_{t+1}$ on lagged values of rational expectations $E_{t-k}x_{t+1}$ and past innovations of $x_{t-k} - E_{t-k-1}x_{t-k}$ for $k \geq 1$. The coefficients on $E_{t-1}x_{t+1}$ and $x_{t-1} - E_{t-2}x_{t-1}$ are in theory equal to $1-\lambda$ and γ. In column 6, we report regression results for the baseline condition as in columns 1 and 3. In column 7, we report regression results for the condition where participants are only required to forecast F_tx_{t+1} – thus not the short-term expectation. We test equality of coefficients on the first lags in the bottom panel. Each subject has a different draw of the process, so we cluster error terms across the two conditions at the subject level – thus allowing for within subject correlation of errors but not across subjects. t-stats are between brackets.

ONLINE APPENDIX

A Additional tests

A.1 Formulation with lags

In this appendix, we explore the following alternative specification of our empirical model of expectation formation.

Firs, note that our recursive specification in equation (6) is equivalent to:

$$F_t x_{t+1} = (1-\lambda)\sum_{k\geq 0}\lambda^k E_{t-k}x_{t+1} + \gamma\sum_{k\geq 0}\lambda^k\left(x_{t-k-1} - E_{t-k-2}x_{t-k-1}\right) \tag{11}$$

Assume that $\gamma = .45$ and $\lambda = .2$ (this corresponds to average values from our main Table 4). In this case, we would expect to have:

$$\begin{aligned} F_t x_{t+1} \approx\ & .8E_t x_{t+1} + .16E_{t-1}x_{t+1} \\ & + .45\left(x_{t-1} - E_{t-2}x_{t-1}\right) + .09\left(x_{t-2} - E_{t-3}x_{t-2}\right) \end{aligned}$$

where we neglects the longer lags.

We thus run a regression using equation (11) and report the results in Table A.1. The coefficient on the first lag hovers between .64 and .8, which is consistent with our baseline results. The coefficient on the second lag (between .26 and .41) is –in some cases significantly - larger than the .16 we are expecting. The two coefficient on one- and two-period lagged extrapolation are very close –and statistically similar– to the .45 and .09 that we expected.

A.2 An alternative model: Adaptive expectations

An alternative formulation of sticky expectations is the traditional notion of adaptive expectations (Nerlove, 1958), which has been used in earlier experimental studies (Dwyer et al. (1993a), Hey (1994a)). Adaptive expectations also contain the notion that expectation formation incorporates new information more slowly than rational expectations, but the recursive formulation differs from specification (6):

$$F_t^i x_{it+1} - x_{it} = \lambda\left(F_{t-1}^i x_{it} - x_{it}\right) + \gamma(x_{it} - E_{t-1}x_{it}) + u_{it+1} \tag{12}$$

Table A.1: Modeling Expectation Formation
Model with lags

	$F_t x_{t+1} - E_t x_{t+1}$			
	(1)	(2)	(3)	(4)
	3 lags	2 lags	$t \leq 20$	$t > 20$
$E_t x_{t+1}$	.69***	.69***	.58***	.79***
	(8.1)	(8.1)	(5.1)	(7.7)
$E_{t-1} x_{t+1}$	.29**	.39***	.46***	.33***
	(2.5)	(4.2)	(3.6)	(2.9)
$E_{t-2} x_{t+1}$	.094			
	(1.2)			
$x_t - E_{t-1} x_t$	.49***	.49***	.51***	.47***
	(12)	(12)	(9.1)	(9)
$x_{t-1} - E_{t-2} x_{t-1}$	.053**	.035*	.047*	.022
	(2)	(1.8)	(1.9)	(.91)
$x_{t-2} - E_{t-3} x_{t-2}$	.027			
	(1.6)			
N	6179	6346	3006	3340
r2	.45	.45	.45	.46

Note: On the panel of participants for which $\rho \in \{0, .2, .4, .6, .8\}$, we run the following regression:

$$F_t x_{t+1} = (1-\lambda) \sum_{k=0}^{n} \lambda^k E_{t-k} x_{t+1} + \gamma \sum_{k=0}^{n} \lambda^k \left(x_{t-k} - E_{t-k-1} x_{t-k} \right)$$

In column 1, we estimate the model assuming $n = 2$; in column 2, we stop at $n = 1$. In columns 3 and 4 we split the sample between the first and last 20 rounds of testing.

where this specification differs from our main specification in two respects. First, the benchmark with which we compare the forecast is the *past* realization of the signal x_{it} instead of the rational expectation about the future signal $E_t x_{it+1}$. These two formulations are equivalent only when $\rho = 1$, i.e. when the process is a random walk. Second, the past expectation component is not the past forecast of x_{it+1}, $F^i_{t-1}x_{it+1}$, but the past forecast over x_{it}, $F^i_{t-1}x_{it}$. Hence, the adaptive formulation does not make use of the term structure of expectations that our main specification exploits. Overall, this approach does *not* nest rational expectations as a particular case. Given that we can vary ρ, we are able to easily distinguish our formulation from the above adaptive-extrapolative model. To do this, we run regression (12) separately for each value of ρ, and ask whether the results are stable across specification.

We report the results in Table A.2, which exactly replicates Table 6 with the adaptive-extrapolative model. Clearly, the estimates of adaptiveness λ and extrapolation γ are quite unstable across values of ρ. For $\rho \in \{.6, .8\}$, none of the parameters is statistically significant. For all conditions except when $\rho = 1$ (in which case the new model is very close to our main specification), there is no trace of extrapolation, as γ is either negative, or negligible or insignificant, and in any case unstable. Similarly, the adaptiveness coefficient is positive and significant for $\rho = 0, .2, 1$ but differs widely. It is insignificant for the other values of persistence. It looks like incorporating rational expectation ρx_{it-1} instead of past realizations x_{it} has the virtue of stabilizing the model.

Table A.2: Adaptive-Extrapolative Model
Sample split by value of ρ

	$F_t x_{t+1} - x_t$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\rho =$	0	.2	.4	.6	.8	1
$F_{t-1}x_t - x_t$	.099**	.093**	-.0068	.12	.071	.53***
	(2.3)	(2.2)	(-.21)	(1.5)	(1.3)	(3.9)
$x_t - E_{t-1}x_t$	-.45***	-.27***	-.19***	.0074	.083	.7***
	(-6.7)	(-5.7)	(-4.7)	(.13)	(1.5)	(4.8)
N	1216	1216	1368	1482	1064	1520
r2	.22	.12	.027	.025	.0052	.3

Note: On the panel of participants for which $\rho \in \{0, .2, .4, .6, .8, 1\}$, we run the following regression:

$$F_t^i x_{it+1} - x_{it} = \lambda \left(F_{t-1}^i x_{it} - x_{it} \right) + \gamma (x_{it} - E_{t-1} x_{it}) + u_{it+1}$$

This model is the adaptive-extrapolative formulation in equation (12). In each column, we estimate the above equation for all participants with a given value of ρ. t-stats between brackets.

A.3 Estimate robustness across Realizations of the process

In this Appendix, we use the experiment where subjects are randomly sorted into 10 conditions. In each condition, the persistence parameter is $\rho = .6$. Within each condition, the path of realized innovations ϵ_{it} is the same for all participants, but it differs across conditions.

For each condition c separately, we run our main specification:

$$F_t^i x_{it+1} - \hat{E}_t x_{it+1} = \lambda_c \left(F_{t-1}^i x_{it+1} - \hat{E}_t^i x_{it+1} \right) + \gamma_c (x_{it} - \hat{E}_{t-1}^i x_{it}) + u_{it+1}$$

where we use the LS rational expectation $\hat{E}$. We then report these estimates in each column of Table A.3. Parameters are consistent with our main results in Table 4 and reasonably consistent with one another.

Table A.3: Modeling Expectation Formation
Model with lags

	$F_t x_{t+1} - E_t x_{t+1}$									
Realization	1	2	3	4	5	6	7	8	9	10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$F_{t-1}x_{t+1} - E_t x_{t+1}$	.18***	.17***	.18***	.27***	.22***	.21***	.19**	.19***	.005	.45***
	(6.7)	(5.8)	(5.7)	(4.3)	(5.3)	(5.9)	(2.1)	(4.4)	(.072)	(3.1)
$x_t - E_{t-1}x_t$	.41***	.64***	.42***	.54***	.46***	.31***	.32***	.46***	.13	.54***
	(9.3)	(15)	(7.5)	(7.7)	(8.8)	(9.1)	(3.6)	(7.6)	(1.2)	(8.1)
N	1369	1184	1369	1110	1184	1221	999	1221	962	1591
r2	.13	.35	.092	.18	.16	.1	.079	.16	.012	.27

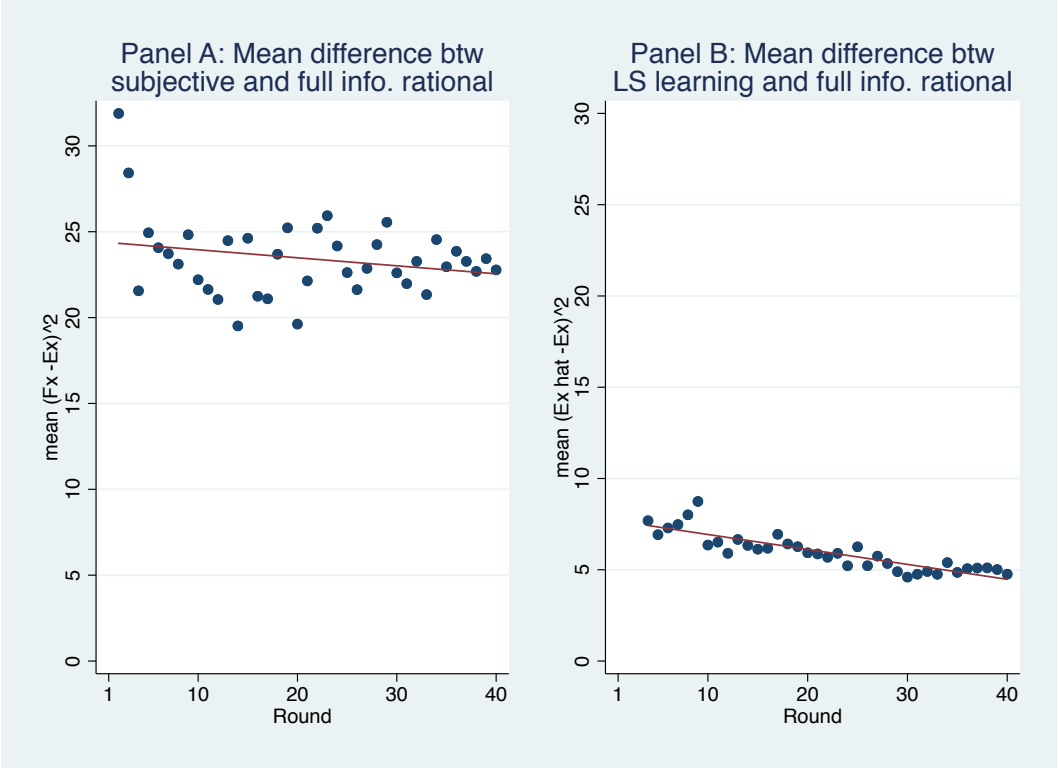
Note: There is only one process with $\rho = .6$. Subjects are randomly allocated to 10 different conditions where there is a single realization of innovation draws ϵ. Thus, within each conditions, all subjects forecast the "same" variable. For each condition c separately, we run our main specification:

$$F_t^i x_{it+1} - \hat{E}_t x_{it+1} = \lambda_c \left(F_{t-1}^i x_{it+1} - \hat{E}_t^i x_{it+1} \right) + \gamma_c (x_{it} - \hat{E}_{t-1}^i x_{it}) + u_{it+1}$$

where we use the LS rational expectation $\hat{E}$ instead of the full information rational expectation. We then report these estimates in each column of the table.

B Appendix Figures

Figure B.1: Mean Square Error in Least Square Learning as a Function of Time



Note: Each period, for each participant, we compute the square of the difference $\left(E^{FI}_{t-1}x_t - E^{LS}_{t-1}x_t\right)^2$. We then take the average of this squared difference across participants, and plot it against time.

C Survey Appendix

C.1 Sample Experiment

Below are the instructions for a sample experiment (Experiment 1, $rho = 0.6$, $\mu = 0$, $sigma = 20$). Participants first see a consent form with brief descriptions of the study. Once they agree to the consent, they will proceed to experimental instructions. The experiment starts with the forecasting task and is then followed by demographic questions. The demographic questions are the same for all of our experiments. The forecasting task may differ slightly depending on the treatment condition, as described in Section 4. We discuss these variants in the next subsection.

Consent Form

Purpose of research: The purpose of this research is to study how people make predictions.

What you will do in this research: You will make forecasts about future realizations of a random process on a web-based platform, followed by a few demographics questions. There are 40 rounds, and you will make 2 predictions per round. You may exit the platform at any time or skip some questions without penalty.

Time required: It takes about 20 minutes to complete the study. You are free to spend as much time as you like up to 60 minutes.

Risks: There are no anticipated risks associated with participating in this study.

Compensation: You will receive **base payment** of **$1.80**. You will also receive a **bonus payment**. The **bonus payment** will be on the scale of **$2.50**, but the precise amount will depend on the accuracy of your predictions.

Your **base payment** and **bonus payment** will be distributed together within one week via MTurk.

Please feel free to contact us with the contact information below or through MTurk if you have any questions about payments. A summary of your payments will be displayed at the end of the study. You may save that page for your records.

Confidentiality: The system allows us to see MTurk Worker IDs and IP addresses. We may use these information for handling payments and to verify data quality, for example that you are in the United States and have not taken our previous surveys. Please make sure to mark your Amazon Profile as private if you do not want it to be found from your MTurk Worker ID. If you communicate with us via email to discuss any issue related to your participation, we will keep your information confidential. All personally identifiable information will be handled in compliance with Harvard and MIT data security requirements, will not be accessible to anyone outside the study team, and will not be used in our data analysis. Data analysis will be based on de-identified data. Part or all of the de-identified data may be shared with other researchers or be made available publicly for academic replication after publication.

Benefits: Your input will help our research develop a better understanding about how people make forecasts. We appreciate your participation. We hope you will also find the survey questions to be interesting.

Contact: If you have any questions, concerns, or suggestions related to this study, the researcher can be reached at:

David Thesmar Sloan School of Management, Massachusetts Institute of Technology 30 Memorial Dr, Cambridge, MA 02142 Cambridge, MA 02139 Email: thesmar@mit.edu (617) 324-7023

This research has been reviewed by the Committee on the Use of Human Subjects in Research at MIT. They can be reached at 617-253-6787, 77 Massachusetts Avenue, Room E25-143B, Cambridge, MA 02139, or couhes@mit.edu. You can contact them for any of the following:

- If your questions, concerns, or complaints are not being answered by the research team,
- If you cannot reach the research team,
- If you want to talk to someone besides the research team, or
- If you have questions about your rights as a research participant.

Please print or screenshot this page for your records.

By selecting to continue, you indicate that you are at least 18 years old and you agree to complete this HIT voluntarily.

[I Give My Consent]

(page break)

Experimental Instructions

Thank you very much for your participation. This study will take you about 20 minutes to complete.

You will receive base payment of **$1.80**. You will also receive a **bonus payment**. The typical bonus amount will be around **$2.50**, but the precise amount will depend on the accuracy of your predictions.

In this study, we would like to understand how people make predictions about future realizations of random processes. We will first show you 40 past realizations of a process, and you will make predictions of its future value for 40 rounds.

You will receive a score for each prediction you make. The more accurate your predictions are, the higher your score will be. If your prediction is out of a certain neighborhood around the actual value, you may receive a score of zero. The specific formula for the score of each prediction is $100 \times \max\{0, 1 - |\Delta|/20\}$ where Δ is the difference between your prediction and the realized value. We estimate that the best performer will receive an average score of 36 per prediction.

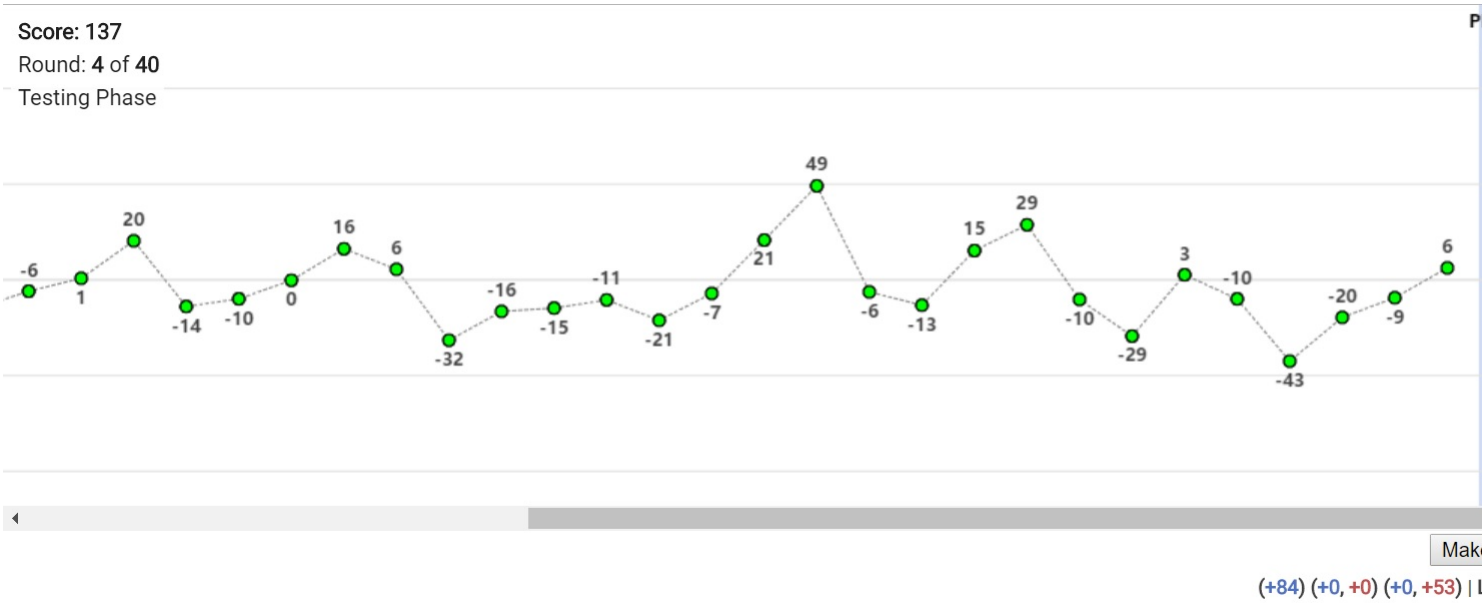
In each of the 40 rounds, we will ask you to predict the next two values of the process. At the end of the experiment, we will calculate your total score in the 40 rounds of predictions. *You will receive the bonus payment in U.S. dollars which is equal to your total score divided by 600.*

[Start Experiment]

(page break)

(This plot is a screenshot of the interactive experimental interface. The green dots indicate past realizations of the statistical process. In each round t, participants are asked to make predictions about two future realizations $F_t x_{t+1}$ and $F_t x_{t+2}$. They can drag the mouse to indicate $F_t x_{t+1}$ in the purple bar and indicate $F_t x_{t+2}$ in the red bar. Their predictions are shown as yellow dots. The grey dot is the prediction of x_{t+1} from the previous round $F_{t-1} x_{t+1}$; participants can see it but cannot change it.

After they have made their predictions, participants click “Make Predictions" and move on to the next round.



The total score is displayed on the top left corner, and the score associated with each of the past prediction (if the actual is realized) is displayed at the bottom.)

Background Information

The prediction section is now over. We would now like to ask a few questions about yourself to help us in our research.

1. What is your gender?
    - Male
    - Female
2. What is your age?
3. What is the highest level of educational degree that you hold?
    - Graduate school (e.g. Masters, Ph.D., Post-doctoral degrees)
    - College
    - High school
    - Below high school
    - Other:
4. Have you taken statistics classes?
    - Yes
    - No

5. Do you have any experience investing in financial assets (e.g. stocks, bonds, mutual funds, pension funds, etc.)?

   - I have extensive experience investing in financial assets.
   - I have some experience.
   - I have very limited experience.
   - I have no experience at all.

6. What is the median of the following numbers? 10, 30, 60, 70, 90, 150, 220, 760

7. A town has two hospitals. The larger hospital has on average 35 babies born every day. The smaller hospital has on average 10 babies born every day. We know that about 50 percent of babies are boys. For a period of 6 months, the hospitals recorded the number of days when more than 70 percent of the babies born are boys, and called them "baby boy days." Which of the following do you think is most likely?

   - The larger hospital recorded more "baby boy days" than the smaller hospital.
   - The smaller hospital recorded more "baby boy days" than the larger hospital.
   - The two hospitals recorded the same number of "baby boy days."

8. A fair coin is tossed 6 times. What do you think about the likelihood of seeing Pattern A: H-T-H-T-T-H vs. Pattern B: H-H-H-T-T-T?

   - Pattern A is more likely than Pattern B
   - Pattern B is more likely than Pattern A
   - They are equally likely
   - None of the above

9. When would you say is a good time to invest in stocks:

   - If the stock market has been going up in the past two years
   - If the stock market has been going down in the past two years
   - I do not have an opinion

Feedback

The study is now completed. Do you have any comments and suggestions for the survey? Did you find anything to be unclear or confusing?

Submit Results

Click the button below to validate and submit your experiment data. This button will submit your HIT for approval and return you to Mechanical Turk.

[Submit Results]

(page break) **Almost done!**

The experiment is now completed. Thank you very much for your participation!

Your total score in the prediction section was [ ].

Base payment: [ ]

Bonus payment: [ ]

You will receive your payments within five days. Bonus payments may vary by +/- one cent due to rounding. Make sure to save this page for your records. If you have any questions, please feel free to contact us.

More Information

In case you are curious about the statistical questions at the end of the experiment, here are the answers. Your answers to these questions do not affect your payments or the quality of your performance in this HIT.

Q. What is the median of the following numbers: 10, 30, 60, 70, 90, 150, 220, 760?

A: The median is (70 + 90) / 2 = 80.

Q. A town has two hospitals. The larger hospital has on average 35 babies born every day. The smaller hospital has on average 10 babies born every day. We know that about 50 percent of babies are boys. For a period of 6 months, the hospitals recorded the number of days when more than 70 percent of the babies born are boys, and called them "baby boy days." Which of the following do you think is most likely?

A: The smaller hospital recorded more "baby boy days" than the larger hospital.

Q. A fair coin is tossed 6 times. What do you think about the likelihood of seeing Pattern A: H-T-H-T-T-H vs. Pattern B: H-H-H-T-T-T?

A: They are equally likely.

To help us with our research, please do not discuss or share these questions on public forums. Thank you very much for your cooperation!

[Submit HIT and Return to MTurk]

C.2 Variants

All experimental conditions in Experiment 1 and Experiment 2 described in Section 4 follow the sample experiment above, except they vary in the parameter ρ.

Several experimental conditions in Experiment 3 have some slight differences, which are explained below.

- Experiment C1 to C4 (context):
  - In third paragraph of experimental instructions, we explain the following

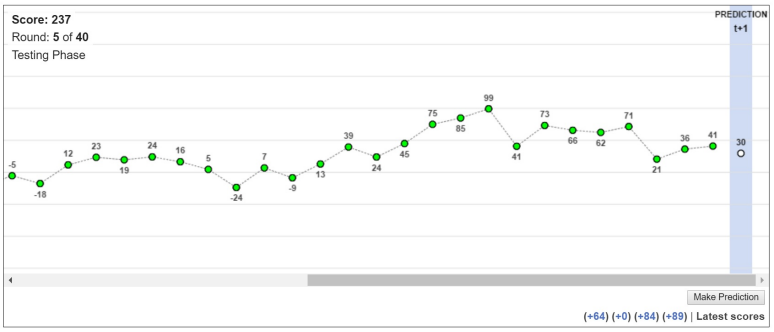
    "In this study, we would like to understand how people make predictions about future realizations of random processes. **The process you will see has the same property as quarterly real GDP growth/monthly inflation/monthly stock returns/monthly house price growth in the US in the last three decades.** We will first show you 40 past realizations of a process, and you will make predictions of its future value for 40 rounds."
  - The parameters rho, μ, σ are based on the properties of these actual processes.

- Everything else is the same as the sample experiment above.

- Experiment C5 to C8 (no context, comparison):
  - The parameters rho, μ, σ correspond to those in Experiments C1 to C4.
  - Everything else is the same as the sample experiment above.

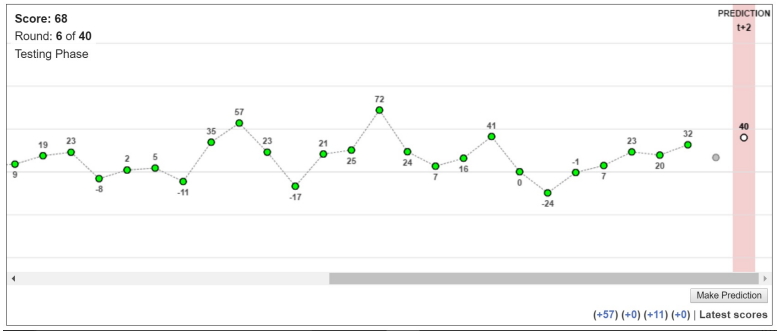
- Experiment C9 (comparison:
  - Everything else is the same as the sample experiment above. $\rho = 0.2$.

- Experiment C10 to C13 (forecast next realization F1 only):
  - Only forecast the next realization (instead of the next two realizations. Below is a screenshot.



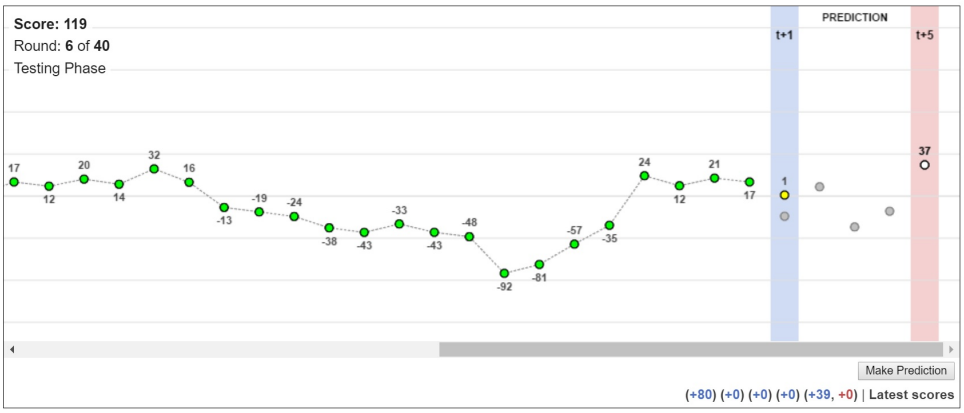
  - Everything else is the same as the sample experiment above.

- Experiment C14 to C17 (forecast two step ahead realization F2 only):
  - Only forecast the two step ahead realization (instead of the next two realizations. Below is a screenshot.



  - Everything else is the same as the sample experiment above.

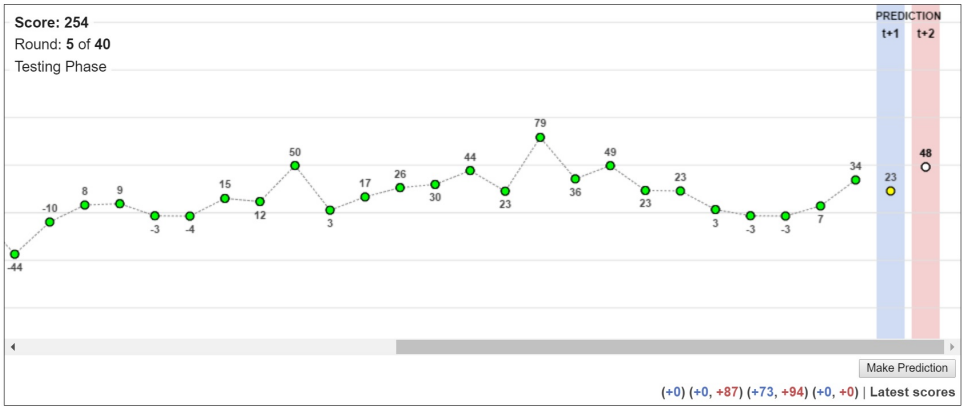
- Experiment C18 to C21 (forecast F1 and F5):



  - Forecast the next realization and the five step ahead realization. Below is a screenshot.
  - Everything else is the same as the sample experiment above.

- Experiment C22 to C25 (no gray dot):

  - Forecast the next two realizations, but remove the gray dot indicating $F_{t-1}x_{t+1}$. Below is a screenshot.



  - Everything else is the same as the sample experiment above.