Analysis Plan for "Complexity and Under- vs. Overreaction in Expectation Formation"

Simin He* Simas Kučinskas†

June 15, 2019

Abstract

We experimentally study how under- and overreaction to new information is affected by *complexity*. Our hypothesis is that people are more likely to underreact to news when information is complex and difficult to process, leading to context dependence in expectation formation. In our experiment, subjects predict future values of variable A. In the *simple treatment*, A follows an AR(1), and subjects only observe past values of A. In the *complex treatment*, subjects additionally observe a leading indicator B, with A and B jointly generated by a bivariate VAR(1). Our experimental design ensures that the predictability and persistence of A are kept constant across treatments. We investigate how under- and overreaction to new information varies with complexity.

Keywords: complexity; overreaction; underreaction; expectation formation.
JEL Classification: C53, C93, D83, D84.

*Shanghai University of Finance and Economics (he.simin@mail.shufe.edu.cn, hesimin.com).
†Humboldt University of Berlin (simas.kucinskas@hu-berlin.de, simaskucinskas.com).

1 Introduction

Do people under- or overreact to news when forming expectations? The answer to this basic question has important consequences for economics and finance. In macro, for example, *under*reaction to new information can explain why monetary policy has real effects.[1] In finance, *over*reaction can generate phenomena such as stock market bubbles and crashes.[2] However, the two views on expectations—under- vs. overreaction—do not square easily with each other. Existing empirical studies are also inconclusive, with evidence for both under- and overreaction.[3]

In this project we experimentally investigate whether under- and overreaction is context dependent and causally affected by *complexity*. With context dependence, people may underreact to new information when forming expectations about inflation yet overreact when thinking about the stock market. Our focus on complexity is theoretically motivated by the fact that many existing models of underreaction are based on the idea that collecting, processing, and using information is costly (Stigler, 1961; Gabaix and Laibson, 2001; Mankiw and Reis, 2002; Sims, 2003; Woodford, 2003; Gabaix, 2014, among many others). Hence, underreaction would seem more likely to occur in complex environments, as in such environments processing information is more difficult. For example, forecasting inflation accurately requires incorporating multiple pieces of information, including changes in monetary policy, oil price developments, technology shocks, and so on. Hence, a plausible reason for observing underreaction in inflation expectations is the complexity of the forecasting task.

Finding good proxies for complexity in the field, let alone sources of exogenous variation for it, is challenging.[4] To overcome this challenge, we conduct a large-scale experiment. We induce complexity by carefully manipulating the difficulty of the forecasting task. In the experiment, subjects predict future values of variable A. In the *simple treatment*, subjects only observe past values of A, and A follows a univariate AR(1) process. In the *complex treatment*, subjects also observe a leading indicator B, with A and B generated by a bivariate VAR(1). Our experimental design ensures that

[1] See, among others, Lucas (1973), Woodford (2003), and Maćkowiak and Wiederholt (2009).

[2] See, among others, Cutler, Poterba, and Summers (1990), DeLong, Shleifer, Summers, and Waldmann (1990), and Barberis, Greenwood, Jin, and Shleifer (2018).

[3] For example, Coibion and Gorodnichenko (2012, 2015) document that survey expectations of macroeconomic variables exhibit underreaction. Vissing-Jorgensen (2004) and Greenwood and Shleifer (2014) show that survey expectations of stock market returns are extrapolative, suggesting overreaction. See also Fuhrer (2017), Bordalo, Gennaioli, Ma, and Shleifer (2018a), and Ryngaert (2018), among many others, as well as the extensive surveys by Pesaran and Weale (2006, Section 5), Manski (2018), Coibion, Gorodnichenko, and Kamdar (2018) and Gennaioli and Shleifer (2018).

[4] Omitted variables bias is an important concern in the field. For example, we may observe underreaction in some settings rather than others because of differences in strategic incentives (Ottaviani and Norman, 2006; Marinovic, Ottaviani, and Sorensen, 2013).

the predictability of A as well as its univariate autocorrelation function is kept constant as we vary complexity. Keeping the univariate autocorrelation function of A constant is not trivial but key for having an apples-to-apples comparison across treatments. Given our design, the only difference in the complex treatment is that the subjects need to combine information from two sources rather than one. Our complex treatment constitutes a minimal deviation from the simple treatment. We only add one new variable and introduce no exogenous costs to gathering information. In that sense, our design is conservative and likely provides a lower bound on the effects of complexity in reality.

To guide our analysis, we first develop a simple model of expectation formation. In the model, subjects may misperceive the persistence of A and B as well as the degree to which A and B are correlated. While simple, the model nests both full-information rational expectations as well as a variety of potential alternatives. We show that if subjects underestimate the degree to which A and B are correlated, they may exhibit more underreaction in the complex treatment. However, the model also highlights that underreacting to B may simultaneously lead people to overreact *more* to A (relative to the simple treatment). Intuitively, if subjects think that B is less important for predicting A, their forecasts load up more on past values of A. However, ignoring B leads the subjects to think that past values of A are more important for prediction than they actually are. All in all, even if people partially neglect B, whether or not underreaction occurs in the complex treatment is an empirical question—a question that we tackle experimentally.

Complexity is not the only reason why reaction to news may exhibit context dependence. Gabaix (2017, Section 2.3.13) develops a model in which overreaction is more likely when the variable being predicted is less persistent. Intuitively, if people exhibit limited attention and anchor to some default level of persistence, they are more likely to overreact when the variable being predicted is less persistent. This insight is a key reason for why we work hard to keep the univariate properties of A, including its persistence, constant across treatments. Had we not done so, complexity would be confounded with the limited-attention-to-persistence channel emphasized by Gabaix. That said, the two channels are complementary and likely operate together in reality. For example, inflation is both more persistent than stock returns and more predictable by other variables. In addition, the two channels both seem to be driven by the same psychological mechanisms, including limited attention and cognitive constraints.

2 Experimental Design

Subjects make one-step-ahead predictions of a time-series variable A_t. Subjects are rewarded for the accuracy of their predictions. Following Dwyer, Williams, Battalio, and Mason (1993) and Landier, Ma, and Thesmar (2019), we use a linear scoring rule which is bounded by zero. For each prediction, subjects receive a score S calculated by

$$S = 100 \cdot \max\left\{0, 1 - |e|/\sigma_\varepsilon\right\},$$

where e is the forecast error (realized value of A_t – forecast of A_t), and σ_ε is the standard deviation of shocks to A_t. We cumulate the scores received in each round and convert the total score at the end of the experiment to dollars using a conversion rate of 500 points = 1 dollar.

Throughout the paper, we refer to forecasts that maximize the expected score for each period as *optimal forecasts*. Given our assumption below that shocks are normally distributed, the optimal forecast of A_{t+1} is the conditional expectation $\mathbb{E}_t[A_{t+1}]$. These optimal forecasts are also equal to full-information rational expectations (Muth, 1961).

2.1 Data-Generating Processes

2.1.1 Simple Treatment

In the *simple treatment*, the variable being predicted follows an AR(1):

$$A_t = \mu(1-\phi) + \phi A_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\varepsilon^2). \tag{1}$$

Here, $\phi \in (-1, 1)$ denotes the *persistence* of A_t, and μ is the unconditional mean, $\mathbb{E}[A_t]$. The variance of the shocks is given by $\sigma_\varepsilon^2 > 0$. For optimal one-step-ahead forecasts, σ_ε^2 is also the mean-squared forecast error. The initial value for the recursion is drawn from the unconditional distribution of A_t.

2.1.2 Complex Treatment

In the *complex treatment*, subjects also predict A_t. However, they additionally observe a second variable B_t. The two variables are generated by a bivariate VAR(1):

$$\begin{aligned} A_t &= \mu(1-\phi_1-\phi_2) + \phi_1 A_{t-1} + \phi_2 B_{t-1} + \varepsilon_t \\ B_t &= \mu(1-\phi) \qquad\qquad\qquad\qquad + \phi B_{t-1} \;\; + \eta_t \end{aligned} \tag{2}$$

The shocks follow $(\varepsilon_t, \eta_t)^\top \sim \text{i.i.d. } \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.

We parametrize the process as

$$\begin{aligned}
\phi_1 &= \frac{\phi(1-\rho^2)}{1-\phi^2\rho^2} \\
\phi_2 &= \frac{\rho(1-\phi^2)}{1-\phi^2\rho^2} \\
\mathbf{\Sigma} &= \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \frac{\sigma_\varepsilon^2(1-\phi^2\rho^2)}{1-\rho^2} \end{pmatrix}
\end{aligned} \tag{3}$$

In this parametrization, $\phi \in (-1, 1)$ measures the *persistence* of A_t and B_t, while $\rho \in (-1, 1)$ is the *correlation* between A_t and B_{t-1}. Hence, ρ captures the extent to which A_t is predictable from past values of B_t. Going back to the example in the introduction, we may think of forecasting inflation as a setting with high ϕ and ρ, while predicting stock returns would correspond to low (i.e., close to zero) values for both of these parameters. When $\phi > \rho$, past values of A_t are *more important* when predicting A_t, while past values of B_t are more important when $\phi < \rho$.[5] We initialize the process using the stationary distribution of A_t and B_t.

We obtain the process in Eqs. (2) and (3) by looking for a bivariate VAR(1) process with the following desired properties:

- P1 (Constant predictability.) Predictability of A_t is the same as in the simple treatment;
- P2 (Constant univariate properties of A_t.) The univariate autocorrelation function of A_t is the same as in the simple treatment;
- P3 (Symmetry.) A_t and B_t have the same univariate autocovariance function.

2.2 Experimental Procedures

2.2.1 Treatments

In a between-subjects design, subjects are assigned to either the simple or complex treatment. In the simple treatment, subjects observe past values of A_t, and are asked to predict future values of A_t. In the complex treatment, subjects observe past values of both A_t and B_t, and are asked to predict future values of A_t.

In addition to information complexity, we also vary the persistence of A_t and B_t (ϕ) as well as the degree to which A_t and B_{t-1} are correlated (ρ). This way, we can study how the effect of information complexity depends on the informativeness of A_t and B_t.

[5] Here, we mean that the optimal forecast of A_{t+1} that only uses A_t has a lower mean-squared error than the optimal forecast that only uses B_t if and only if $\phi > \rho$.

Table 1

Treatments Summary

Notes: The table summarizes the treatments that we run experimentally. We vary the information context, persistence, and correlation, as shown in the table, with six treatment arms in total. In all treatments, we set $\mu = 100$ and $\sigma_\varepsilon = 10$.

Treatments:	Simple	Complex	
		Low correlation ($\rho = 0.25$)	High correlation ($\rho = 0.75$)
Low persistence ($\phi = 0.25$)	Simple-1	Complex-1	Complex-3
High persistence ($\phi = 0.75$)	Simple-2	Complex-2	Complex-4

In total, we have the following treatment design. There are three types of variations: information context (simple or complex), persistence (low or high), and the correlation between A_t and B_{t-1} in the complex treatments (low or high). This yields six treatment arms in total. In all treatments, we set the expected value of the variables to $\mu = 100$ and use a standard deviation of $\sigma_\varepsilon = 10$ for the shocks to A_t. The treatment summary can be seen in Table 1.

As can be seen from Table 1, there are in total six treatment arms, two simple and four complex. In the simple treatments, the persistence ϕ is either low (0.25) or high (0.75). In the complex treatments, the persistence ϕ is either low (0.25) or high (0.75), and the correlation ρ is either low (0.25) or high (0.75).

2.2.2 Subjects Recruitment

We use Amazon's Mechanical Turk (AMT) to recruit subjects and conduct our experiment. AMT is an online labor market that is commonly used in social sciences (e.g., Horton, Rand, and Zeckhauser, 2011; Cavallo and Cruces, 2015; Kuziemko, Norton, Saez, and Stantcheva, 2015; DellaVigna and Pope, 2018).

We plan to recruit about 1,000 subjects (called workers on AMT). We will first create one Human Intelligence Task (HIT) with a recruiting quota of 1,000 workers, and this HIT will be valid for 2 weeks. If we fail to recruit 1,000 workers in this wave, we will post a second HIT after the first HIT expires. Reposting the task can potentially attract more workers, as the new HIT will appear on the top of the available tasks on AMT. The reposted HIT will also be valid for 2 weeks. We will continue until we get about 1,000 workers. Across all the HITs we post for this experiment, we will not allow any workers to participate for more than once.

When recruiting, we use block randomization to randomly allocate each worker into one of the six treatment arms. In principle, this allows the 1,000 workers to be perfectly equally distributed across the 6 treatments. However, in practice, workers may drop out and not complete the experiment they sign up for. In this case, AMT recruits new workers until 1,000 individuals complete our experiment. If the dropout rate is different across treatments, we may end up recruiting imbalanced numbers of workers across treatments, but we do not expect a big difference.

In order to increase the quality of data on AMT, we impose the following requirements when recruiting workers: (1) we restrict all workers to be from the USA, (2) we restrict the workers to have a success rate of 95% or higher, and (3) the workers need to have completed at least 500 HITs.

2.2.3 Experimental Protocol

An AMT worker who sees our posted HIT can preview our experiment on the online platform.[6] In the preview, the worker learns the expected length (12 minutes) and expected earnings ($1.00 base payment, around $2.00 bonus payment) of the experiment. After the preview, a worker who meets our selection criteria can accept to participate. Once the worker accepts, a unique user ID is assigned to this worker. This user ID determines the treatment the worker is assigned to. After accepting the experiment, the worker has 60 minutes to finish the entire experiment.

During the experiment, the workers first read the instructions about the experiment. Then, they proceed to the prediction page. In the prediction page, they first see the initial 40 values of A_t (and B_t in the complex treatments). They are asked to make predictions of A_t for the next 40 rounds. After they make each prediction, they are shown the actual realization of A_t in that round (as well as the realization of B_t in the complex treatments), and their score for that round is cumulated to their total score. The experimental screen is shown in Figure 1. After finishing the 40 predictions, they fill out a short questionnaire. Finally, their total score and total payment (in dollars) is shown.

The worker is paid only if the experiment is fully completed. If a worker fails to complete the entire experiment within 60 minutes after accepting the task, AMT attempts to recruit another worker. The HIT will be available until 1,000 different workers complete the task or the HIT expires. AMT ensures that a worker cannot participate in the same HIT more than once.

[6] The full experimental instructions are provided in Appendix A.

Figure 1

Screenshot of Prediction Page: Complex Treatment

Notes: Screenshot of the experimental screen in the complex treatment. The only difference in the simple treatment is that there is no variable B that is shown. Subjects submit their scores by either clicking on the blue area or typing their prediction in the text box. The experimental screen in the simple treatment is provided in Figure 2.

2.2.4 Pilot

Before conducting our experiment, we did a small pilot study on AMT on May 22, 2019. In the pilot, we created an HIT titled "Prediction task – forecast future values of given variable" and recruited 25 workers. We conducted this pilot to check if there were any technical problems, and to estimate (i) time necessary to finish the task; and (ii) average earnings. The pilot went well without any technical problems. In the instruction of the pilot, we mentioned that the expected length of the experiment was 15 minutes, the base payment was \$1.20, the exchange rate was 400 points per \$1.00, and that we expected the average bonus payment to be around \$2.00. However, it turned out that the average time taken to complete the experiment in fact was about 12 minutes, and we underestimated the average score somewhat. After reevaluation, in the real experiment, we changed our base payment to \$1.00 and the exchange rate to 500 points per \$1.00, which should yield an average bonus payment of around \$2.00.

3 Theoretical Predictions

We now derive the theoretical predictions for over- and underreaction for a simple model of expectation formation. This model nests full-information rational expectations—as well as various deviations from rational expectations—as special cases.

Suppose that while the true parameters governing the data-generating process are given by ϕ and ρ, the subjects perceive these parameters to be equal to ϕ_p and ρ_p; all parameters are assumed to lie in $(0, 1)$. Otherwise, the subjects correctly perceive the structure of the process.[7] Hence, their forecasts are given by

$$\mathbb{F}_t[A_{t+1}] = \begin{cases} \phi_p A_t & \text{in the simple treatment} \\ \phi_{1,p} A_t + \phi_{2,p} B_t & \text{in the complex treatment} \end{cases}$$

Here, $\phi_{1,p}$ and $\phi_{2,p}$ are given by the expressions in Eq. (3) with ϕ and ρ replaced by their perceived counterparts.

A variety of existing models of expectation formation is captured by this simple formulation. In the special case $\phi_p = \phi$ and $\rho_p = \rho$, we recover full-information rational expectations. When $\phi_p > \phi$, subjects perceive the variables to be more persistent than they actually are, whereas with $\phi_p < \phi$ subjects think that they are less persistent. In the simple treatment, underreaction to new information is captured by $\phi_p < \phi$, while

[7] Given our focus on under- and overreaction, it is without loss of generality to assume that subjects correctly perceive μ.

$\phi_p > \phi$ yields overreaction.[8] In the complex treatment, under- and overreaction is driven by both ϕ_p and ρ_p, as characterized below. Partially ignoring the informational content of B_t would be captured by $\rho_p < \rho$, while mistakenly thinking that B_t is more important for predicting A_t than is actually the case is given by $\rho_p > \rho$.

We now derive the predictions of the model for under- and overreaction to new information. To quantify under- and overreaction, we follow the methodology of Kucinskas and Peters (2019) and calculate the theoretically predicted *bias coefficients*. Bias coefficients are direct measures of under- and overreaction. These coefficients are equal to the difference between the perceived response of A_t to some shock to the actual response of A_t to that shock. A positive bias coefficient indicates overreaction, while a negative coefficient means underreaction.

We calculate bias coefficients for the response of expectations to the current shocks, ε_t and η_t. Similarly to Coibion and Gorodnichenko (2012), we normalize the coefficients by the true response of A_t. In the simple treatment, the bias coefficient with respect to ε_t (shock to A_t) is given by

$$b_A^{\text{simple}} = \frac{\phi_p - \phi}{\phi}.$$

In the complex treatment, bias coefficients with respect to ε_t (shock to A_t) and η_t (shock to B_t) are equal to

$$b_A^{\text{complex}} = \frac{\phi_{1,p} - \phi_1}{\phi_1}$$
$$b_B^{\text{complex}} = \frac{\phi_{2,p} - \phi_2}{\phi_2}$$

Finally, we calculate a measure of *overall overreaction*. This measure combines the reaction to all shocks that are present in a given treatment. We obtain this measure from the population regression of forecast errors on the *optimal forecast revision*. The optimal forecast revision is given by

$$\text{rev}_t^* \equiv \mathbb{E}_t[x_{t+1}] - \mathbb{E}_{t-1}[x_{t+1}] = \begin{cases} \phi\varepsilon_t & \text{in the simple treatment} \\ \phi_1\varepsilon_t + \phi_2\eta_t & \text{in the complex treatment} \end{cases}$$

[8] For example, the sticky-information model of Mankiw and Reis (2002) and the noisy-information model of Woodford (2003) both predict underreaction to new information in the AR(1) case. While the pattern of underreaction is not exactly the same as in the model with a misperceived level of persistence, it is qualitatively similar (see Kucinskas and Peters, 2019). By the same logic, extrapolative or diagnostic expectations (Bordalo, Gennaioli, and Shleifer, 2018b) can be approximated by $\phi_p > \phi$.

We obtain our measure of overall overreaction by the slope coefficient β in

$$\mathrm{fe}_{t+1} = \alpha + \beta \mathrm{rev}_t^* + u_{t+1},$$

where forecast errors fe_{t+1} are defined as $\mathrm{fe}_{t+1} = \mathbb{F}_t[A_{t+1}] - A_{t+1}$. With full-information rational expectations, $\beta = 0$. If subjects overreact to new information (as proxied by rev_t^*) and adjust the forecast more than is optimal, $\beta > 0$, while underreaction yields $\beta < 0$. Note that in the simple treatment, β is equal to b_A^{simple}, as it should be given that there is only one shock in the simple treatment. In the complex treatment, β is given by a weighted average of the two shock-specific bias coefficients:

$$\left(\frac{\phi_1^2\sigma_\varepsilon^2}{\phi_1^2\sigma_\varepsilon^2 + \phi_2^2\sigma_\eta^2}\right) b_A^{\text{complex}} + \left(\frac{\phi_2^2\sigma_\eta^2}{\phi_1^2\sigma_\varepsilon^2 + \phi_2^2\sigma_\eta^2}\right) b_B^{\text{complex}}.$$

The weights are given by the fraction of variance in optimal forecast revisions due to each shock.[9] For the model of *diagnostic expectations* (Bordalo, Gennaioli, and Shleifer, 2018b), the coefficient β is equal to the representativeness parameter (θ in the notation of Bordalo, Gennaoili, and Shleifer; c.f. their Proposition 1). This observation provides an additional justification for using β as a measure of overall overreaction.

The following result is immediate from the definitions above.

Proposition 1. *All else equal, overreaction depends on the perceived parameters as follows:*

1. *In the simple treatment, overreaction to ε_t (and hence also overall overreaction) is increasing in the perceived persistence ϕ_p.*
2. *In the complex treatment, overreaction to ε_t is increasing in the perceived persistence ϕ_p and decreasing in the perceived correlation ρ_p. Overreaction to η_t is decreasing in the perceived persistence ϕ_p and increasing in the perceived correlation ρ_p. The effect of changes in ϕ_p and ρ_p on overall overreaction is ambiguous.*

Intuitively, overreaction to the shock to A_t (i.e., ε_t) is more likely when subjects perceive persistence of A_t to be higher or the predictive content of B_t lower. The effects on overreaction to the shock to B_t (i.e., η_t) go in the opposite direction. If subjects underestimate the predictive content of B_t, they are more likely to underreact to η_t. Given

[9] Our measure of overall under- and overreaction is closely related to the composite bias coefficients proposed by Kucinskas and Peters (2019). The key difference is that composite bias coefficients are given by a non-linear function of the shock-specific bias coefficients, rather than a weighted average as here. The key advantage of the measure used in the present paper is its robustness to noise in expectations, as the estimation employs the true shocks.

the opposing effects, the effect of perceived parameters on overall overreaction is ambiguous. In particular, even if people underestimate the importance of B_t in predicting future values of A_t, they may nevertheless overreact to shocks to A_t as well as exhibit overall overreaction. Hence, whether or not people exhibit more underreaction in the complex treatment is an empirical question.

A key insight from the proposition is that the way people react to different sources of information in the complex treatment is interdependent. If people neglect the informational content of B_t more (ρ_p is lower), that increases underreaction to B_t. However, that simultaneously makes the agent overreact more to A_t. The total effect on overall overreaction is ambiguous and depends on the parameter values. This result, while simple, captures an important fact. If subjects have limited attention, focusing too much on one variable must necessarily lead to too little attention allocated to other variables.

We emphasize that the model is reduced form and does not provide a deeper explanation as to how the perceived parameters ϕ_p and ρ_p come about. In particular, we do not explicitly model how complexity affects behavior. In doing so, we follow the existing related experimental literature (see, for example, Enke and Zimmermann, 2019; Enke, 2019). We think of the model as being useful for studying the *proximate* causes of under- and overreaction—which may be either that people misperceive persistence or correlation. This way, the model is helpful for understanding the channels through which complexity may affect expectation formation.[10] The perceived parameters can also be easily estimated empirically, shedding light on the mechanisms through which under- and overreaction occur in the experiment.

4 Data Analysis Plan

In this section, we first describe the sample selection criteria that we use. Then, we describe the key steps of the empirical analysis.

4.1 Data Filtering

Data quality on AMT can be problematic (Ahler, Roush, and Sood, 2019). To minimize the impact of data imperfections on our results, we impose requirements on workers in

[10] A key challenge with explicitly modeling complexity is that in our experiment, complexity stems from costs in processing and using information, while the true informational content is kept constant across treatments. As a result, the standard rational inattention model with entropy-based constraints (Sims, 1998, 2003) is not well suited for our purposes. Generalized rational inattention models (such as those studied by Caplin, Dean, and Leahy, 2017) may be more appropriate, but investigating these possibilities is outside the scope of the present article.

the recruitment stage, as discussed above. In addition, we plan to exclude potentially low-quality data using the following criteria. Specifically, we use the following filters:

1. Exclude if total response time is less than 3 minutes;
2. Exclude if total score is less than 200 points.

For our robustness checks, we plan to do another round of data-quality checks by further excluding subjects according to the following filters:

1. Exclude if do not select "Disagree" for the attention question in the Questionnaire;
2. Exclude if total response time is in the top or bottom 5% of the sample;
3. Exclude duplicated IPs.

4.2 Forecast Accuracy

We first test whether subjects make less accurate forecasts in the complex treatment. Under full-information rational expectations, the score should be equal across the simple and complex treatments. We test this using the non-parametric Mann-Whitney test:

$$H_0 : \text{total score}_{\text{complex}} = \text{total score}_{\text{simple}}.$$

Our conjecture is that the total score (i.e., cumulated score across all rounds of the experiment) is *lower* in the complex information treatment.[11]

4.3 Under- and Overreaction

Our second test is whether there is "more underreaction" in the complex treatment. More precisely, we ask two questions. First, we test whether subjects exhibit more overall underreaction in the complex treatment. Second, we zoom in on the reaction to new information contained in A_t and B_t separately. Our methodology for measuring under- and overreaction follows Kucinskas and Peters (2019) closely.

Given the results in the existing literature and our own pilot study, we hypothesize that any under- or overreaction in the experiment will stem primarily from under- or overreaction to the most recent shocks. That is, when making the prediction at time t, subject i is most likely to under- or overreact to $\varepsilon_{i,t}$ and $\eta_{i,t}$ rather than deeper lags of these shocks. Hence, our regressions will focus on how forecasts react to $\varepsilon_{i,t}$ and $\eta_{i,t}$. Focusing on how forecasts react to the most recent shocks also helps alleviate potential concerns about multiple hypothesis testing.

[11] To be conservative, we use two-tailed tests in all our analysis, e.g., the alternative hypothesis for the present test is $H_A : \text{total score}_{\text{complex}} \neq \text{total score}_{\text{simple}}$.

As in Section 3, forecast errors are defined as $\text{fe}_{i,t+1} = \mathbb{F}_{i,t}[A_{i,t+1}] - A_{i,t+1}$ with $\mathbb{F}_{i,t}[A_{i,t+1}]$ denoting the forecast of subject i at time t. A positive forecast error indicates overprediction, and a negative forecast error indicates underprediction. We define *scaled shocks* by $\tilde{\eta}_{i,t} = \phi_2 \eta_{i,t}$ and

$$\tilde{\varepsilon}_{i,t} = \begin{cases} \phi \varepsilon_{i,t} & \text{in the simple treatment} \\ \phi_1 \varepsilon_{i,t} & \text{in the complex treatment} \end{cases}$$

Intuitively, the scaled shocks measure by how much the optimal forecasts should be adjusted after the realization of the new shocks. Finally, let $\text{rev}^*_{i,t}$ denote the total forecast revision of the optimal forecast, again defined in the same way as in Section 3.[12] We use scaled shocks in all our regressions to take into account the fact that the response of A_t to shocks in either A_t or B_t varies across treatments.

For our first subquestion, we estimate

$$\text{fe}_{i,t+1} = \alpha_i + \beta \text{rev}^*_{i,t} + \gamma(\text{rev}^*_{i,t} \times \text{complex}_i) + u_{i,t}.$$

and test $H_0 : \gamma = 0$. The standard errors are clustered by subject. Under full-information rational expectations, forecast errors are unpredictable by variables available at the time of making the forecast, implying $\beta = \gamma = 0$. Note that in this regression, β measures overall overreaction in the simple treatment, while $\beta + \gamma$ provides overall overreaction in the complex treatment.

For our second subquestion, we estimate

$$\text{fe}_{i,t+1} = \alpha_i + \beta \tilde{\varepsilon}_{i,t} + \gamma(\tilde{\varepsilon}_{i,t} \times \text{complex}_i) + u_{i,t},$$

and test $H_0 : \gamma = 0$. Similarly to before, β measures overreaction to shocks to A_t in the simple treatment, and $\beta + \gamma$ gives the same measure for the complex treatment. Then, we use only data from the complex treatments and estimate

$$\text{fe}_{i,t+1} = \alpha_i + \beta \tilde{\eta}_{i,t} + u_{i,t}$$

to measure overreaction to shocks to B_t.

As discussed by Kucinskas and Peters (2019), more precise estimates of overreaction can be obtained using *ex-ante* forecast errors, defined by $\mathbb{F}_{i,t}[A_{t+1}] - \mathbb{E}_{i,t}[A_{t+1}]$, in place

[12] Regresing forecast errors on the scaled shocks is equivalent to regressing forecast errors on past shocks and normalizing the estimated coefficients by the true impulse responses, as done by Coibion and Gorodnichenko (2012).

of the ex-post forecast errors employed above. The reason is that the time $t+1$ shock does not enter the ex-ante forecast error, thereby increasing statistical power. Hence, we also plan to run the regressions above using ex-ante forecast errors.

4.4 Further Tests

Our further tests fall into two categories. First, we run tests designed to investigate the channel through which forecasts may be less accurate and overreaction less pronounced in the complex treatment. Second, we investigate heterogeneity in the responses across treatments.

4.4.1 Mechanism Questions

Is the complex treatment really complex? We first investigate whether the complex treatment is indeed more cognitively taxing by looking at two direct measures of the complexity of the treatments: (a) prediction times; and (b) slope of the learning curve. For (a), we compare the average prediction times for the last 30 rounds of the experiment across the simple and complex treatments using the Mann-Whitney test.[13] For (b), we calculate the average score in the first twenty (avg_score_start_i) and the last twenty rounds (avg_score_end_i) for each subject i. We then approximate the slope of the learning curve by

$$\text{learning curve}_i = \text{avg\_score\_end}_i - \text{avg\_score\_start}_i$$

and test

$$H_0 : \text{learning curve}_{\text{complex}} = \text{learning curve}_{\text{simple}},$$

using the Mann-Whitney test.

How do subjects perceive the data-generating process? Finally, we study how the subjects perceive the parameters of the data-generating process. For the simple treatment, we estimate

$$\mathbb{F}_{i,t}[A_{i,t+1}] = \alpha_i + \phi_p A_{i,t} + u_{i,t}$$

[13] Subjects in our experiment may not start the experiment right away after accepting the task on MTurk, or submit a few predictions to estimate how long it may take to finish the experiment and then continue working on other tasks. As a result, prediction times in the first few rounds are not very accurate measures of prediction times.

to estimate the perceived persistence ϕ_p, separately for each simple treatment (i.e., for each level of ϕ). For the complex treatment, we first estimate

$$\mathbb{F}_{i,t}[A_{i,t+1}] = \alpha_i + \phi_{1,p}A_{i,t} + \phi_{2,p}B_{i,t} + u_{i,t}$$

to estimate the perceived law of motion for A_t. We then invert the true formulas to obtain the perceived persistence ϕ_p and correlation level ρ_p by

$$\phi_p = \frac{[1 + (\phi_{1,p})^2 - (\phi_{2,p})^2] - \sqrt{[1 + (\phi_{1,p})^2 - (\phi_{2,p})^2]^2 - 4(\phi_{1,p})^2}}{2\phi_{1,p}}$$
$$\rho_p = \frac{[1 - (\phi_{1,p})^2 + (\phi_{2,p})^2] - \sqrt{[1 + (\phi_{1,p})^2 - (\phi_{2,p})^2]^2 - 4(\phi_{1,p})^2}}{2\phi_{2,p}}$$

Again, we perform this exercise separately for each complex treatment (i.e., for each level of ϕ and ρ).

Finally, we perform the same exercise for each individual subject separately to investigate the heterogeneity in the perceived parameters. Specifically, we estimate the two regressions above separately for each subject i, yielding estimates of $\phi_{i,p}$ and $\rho_{i,p}$. We are then interested in the distribution of the perceived relative biases, defined by $(\phi_{i,p} - \phi)/\phi$ and $(\rho_{i,p} - \rho)/\rho$ across the treatments.

4.4.2 Heterogeneity

Forecasting accuracy across treatments. We first calculate the forecast score for every treatment (with varying levels of persistence ϕ and correlation ρ) to investigate whether there are systematic differences with respect to these parameters.

Persistence and overreaction to A. If subjects are more likely to overreact when A is less persistent, we may expect the effect of the complex treatment to be larger when persistence is low. To investigate this possibility, we estimate a regression with a triple interaction term

$$\begin{aligned} \text{fe}_{i,t+1} = \alpha_i + \beta \text{rev}^*_{i,t} + \gamma(\text{rev}^*_{i,t} \times \text{complex}_{i,t}) \\ + \kappa(\text{rev}^*_{i,t} \times \text{complex}_i \times \text{high persistence}_i) + u_{i,t} \end{aligned}$$

and test $H_0 : \kappa = 0$.

Correlation and overreaction to B. Subjects may be more likely to underreact to B when B is more important for predicting A. To investigate whether that is true, we estimate

$$\text{fe}_{i,t+1} = \alpha_i + \beta\tilde{\eta}_{i,t} + \gamma(\tilde{\eta}_{i,t} \times \text{high correlation}_i) + u_{i,t},$$

and test $H_0 : \gamma = 0$.

References

Ahler, D. J., Roush, C. E., Sood, G., 2019. The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. Working paper.

Barberis, N., Greenwood, R., Jin, L., Shleifer, A., 2018. Extrapolation and bubbles. Journal of Financial Economics 129 (2), 203–227.

Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2018a. Overreaction in macroeconomic expectations. Working paper.

Bordalo, P., Gennaioli, N., Shleifer, A., 2018b. Diagnostic expectations and credit cycles. The Journal of Finance 73 (1), 199–227.

Caplin, A., Dean, M., Leahy, J., 2017. Rationally inattentive behavior: Characterizing and generalizing Shannon entropy, 1–51.

Cavallo, A., Cruces, G., 2015. Inflation expectations, learning and supermarket prices evidence from field experiments. International Journal of Production Research 53 (9), 2771–2786.

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

Coibion, O., Gorodnichenko, Y., 2015. Information rigidity and the expectations formation process: A simple framework and new facts. The American Economic Review 105 (8), 2644–2678.

Coibion, O., Gorodnichenko, Y., Kamdar, R., 2018. The formation of expectations, inflation and the Phillips Curve. Journal of Economic Literature, forthcoming.

Cutler, D. M., Poterba, J. M., Summers, L. H., 1990. Speculative dynamics and the role of feedback traders. The American Economic Review: Papers & Proceedings 80 (2), 63–68.

DellaVigna, S., Pope, D., 2018. What motivates effort? Evidence and expert forecasts. Review of Economic Studies 85 (2), 1029–1069.

DeLong, B. J., Shleifer, A., Summers, L. H., Waldmann, R. J., 1990. Positive feedback investment strategies and destabilizing rational speculation. The Journal of Finance 45 (2), 379–395.

Dwyer, G. P., Williams, A. W., Battalio, R. C., Mason, T. I., 1993. Tests of rational expectations in a stark setting. The Economic Journal 103 (418), 586–601.

Enke, B., 2019. What you see is all there is. Working paper.

Enke, B., Zimmermann, F., 2019. Correlation neglect in belief formation. Review of Economic Studies 86 (1), 313–332.

Fuhrer, J., 2017. Intrinsic expectations persistence: Evidence from professional and household survey expectations. Working paper.

Gabaix, X., 2014. A sparsity-based model of bounded rationality. The Quarterly Journal of Economics, 1661–1710.

Gabaix, X., 2017. Behavioral inattention. Working paper.

Gabaix, X., Laibson, D., 2001. The 6D bias and the equity premium puzzle. NBER Macroeconomics Annual 16, 257–312.

Gennaioli, N., Shleifer, A., 2018. A Crisis of Beliefs: Investor Psychology and Financial Fragility. Princeton University Press, Princeton.

Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. Review of Financial Studies 27 (3), 714–746.

Horton, J. J., Rand, D. G., Zeckhauser, R. J., 2011. The online laboratory: Conducting experiments in a real labor market. Experimental Economics 14 (3), 399–425.

Kucinskas, S., Peters, F. S., 2019. Measuring biases in expectation formation. Working paper.

Kuziemko, I., Norton, M. I., Saez, E., Stantcheva, S., 2015. How elastic are preferences for redistribution? Evidence from randomized survey experiments. American Economic Review 105 (4), 1478–1508.

Landier, A., Ma, Y., Thesmar, D., 2019. Biases in expectations: Experimental evidence. Working paper.

Lucas, R. E., 1973. Some international evidence on output-inflation tradeofs. American Economic Review 63 (3), 326–334.

Maćkowiak, B., Wiederholt, M., 2009. Optimal sticky prices under rational inattention. The American Economic Review 99 (3), 769–803.

Mankiw, G. N., Reis, R., 2002. Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips Curve. The Quarterly Journal of Economics 117 (4), 1295–1328.

Manski, C. F., 2018. Survey measurement of probabilistic macroeconomic expectations: Progress and promise. NBER Macroeconomics Annual 32, forthcoming.

Marinovic, I., Ottaviani, M., Sorensen, P., 2013. Forecasters' objectives and strategies. In: Handbook of Economic Forecasting. Vol. 2. Elsevier B.V., Ch. 12, pp. 689–720.

Muth, J. F., 1961. Rational expectations and the theory of price movements. Econometrica 29 (3), 315–335.

Ottaviani, M., Norman, P. S., 2006. The strategy of professional forecasting. Journal of Financial Economics 81, 441–466.

Pesaran, M. H., Weale, M., 2006. Survey expectations. In: Elliott, G., Granger, C. W. J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 1. Elsevier, Amsterdam, Ch. 14, pp. 715–776.

Ryngaert, J., 2018. What do (and don't) forecasters know about U.S. inflation? Working paper.

Sims, C. A., 1998. Stickiness. Carnegie-Rochester Conference Series on Public Policy 49, 317–356.

Sims, C. A., 2003. Implications of rational inattention. Journal of Monetary Economics 50 (3), 665–690.

Stigler, G. J., 1961. The economics of information. The Journal of Political Economy 69 (3), 213–225.

Vissing-Jorgensen, A., 2004. Perspectives on behavioral finance: Does 'irrationality' disappear with wealth? Evidence from expectations and actions. NBER Macroeconomics Annual 18, 139–208.

Woodford, M., 2003. Imperfect common knowledge and the effects of monetary policy. In: Aghion, P., Frydman, R., Stiglitz, J., Woodford, M. (Eds.), Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps. Princeton University Press, Princeton, pp. 25–58.

Appendix A Experimental Instructions

A.1 Preview

Welcome to this experiment on forecasting!

The experiment is about predicting future values of a given variable. For participating in the experiment, you will receive a **base payment** of **$1.00**. You can also receive an additional **bonus payment**, and the amount of the bonus payment depends on your performance. We estimate that the average bonus payment will be around **$2.00**. The task will take around **12 minutes** to complete. To receive the payment (both base and bonus), you are required to finish the full experiment.

A.2 Instructions

Welcome to this experiment! Please read the following instructions carefully.

For participating in the experiment, you will receive a **base payment** of **$1.00**. You can also receive an additional bonus payment, and the amount of the **bonus payment** depends on your performance. We estimate that the average bonus payment will be around **$2.00**. The task will take around **12 minutes** to complete. To receive the payment (both base and bonus), you are required to finish the full experiment.

TASK DESCRIPTION

In this experiment, you are going to **predict** future values of **"Variable A"**.

(Only in the simple treatment.) The experiment lasts for 40 rounds. In each round, you will see past values of "Variable A" and "Variable B". Past values of "Variable A" are related to future values of "Variable A", and the relationship is stable. At the beginning of the experiment, you will see data from 40 previous rounds. Then, you will have to make predictions for 40 rounds.

(Only in the complex treatment.) The experiment lasts for 40 rounds. In each round, you will see past values of "Variable A" and "Variable B". Past values of "Variable A" and "Variable B" are related to future values of "Variable A", and the relationship is stable. At the beginning of the experiment, you will see data from 40 previous rounds. Then, you will have to make predictions for 40 rounds.

BONUS PAYMENT

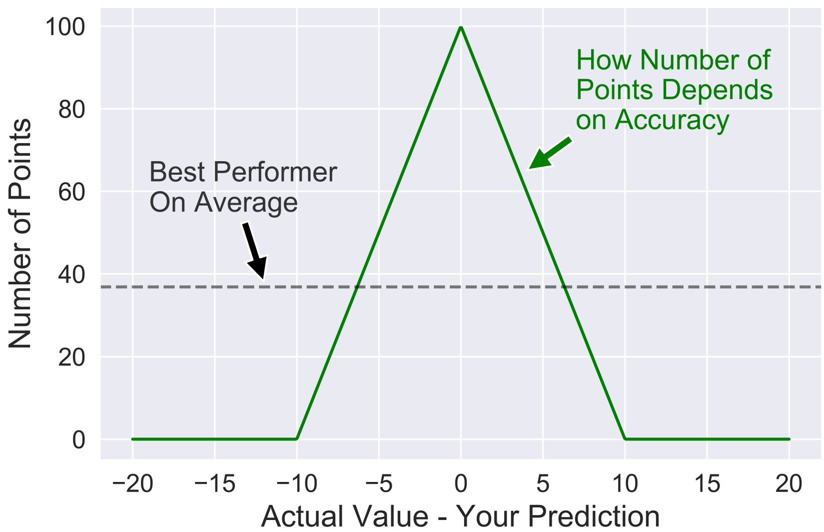
You receive points for accurate predictions. At the end of the experiment, your points will be converted to dollars at the rate of **500 points = $1**. Your total earnings in dollars will therefore be **$1.00 + your total points/500**.

How many points you receive depends on the accuracy of your prediction:

- The closer your prediction is to the actual value, the more points you receive;
- If your prediction is more than 10 units away from the actual value, you receive no points.

We estimate that the best performer will on average receive 37 points per round.

Graphically, the number of points you receive depends on accuracy as follows:



The exact formula for the number of points is $100 * \max\{0, 1 - |D|/10\}$ where D is the difference between the actual value and your prediction.

A.3 Questionnaire

Questions marked with an asterisk are compulsory. The possible answers we provided in closed-ended questions are given in the parentheses.

1. *Age
2. *Gender (Male, female)
3. *Have you ever taken a class on statistics or forecasting? (Yes, no)
4. *What is the highest level of educational degree that you hold? (Below high school, high school, college, graduate school, other)
5. *I am someone who finds it easy to concentrate and can work on tasks for a long time. (Completely agree, agree, neutral, disagree, strongly disagree)
6. *I have a good eye for detail and often notice things that others miss. (Completely agree, agree, neutral, disagree, strongly disagree)
7. *Please select "disagree" among the following options to show that you are paying attention. (Completely agree, agree, neutral, disagree, strongly disagree)
8. *In the experiment, I found past values of "Variable A" to be useful when making predictions. (Completely agree, agree, neutral, disagree, strongly disagree)
9. *(Only in the complex treatment) In the experiment, I found past values of "Variable B" to be useful when making predictions. (Completely agree, agree, neutral, disagree, strongly disagree)
10. Do you have any additional comments about the experiment?

Figure 2

Screenshot of Prediction Page: Simple Treatment

Notes: Screenshot of the experimental screen in the simple treatment. Subjects submit their scores by either clicking on the blue area or typing their prediction in the text box.