Adaptive Behavior in the Lab[*]

Myrna Hennequin[a], Cars Hommes[b,c], Anita Kopányi-Peuker[d], and Eva Levelt[c]

[a]University of Essex, [b]Bank of Canada, [c]University of Amsterdam, [d]Radboud University Nijmegen

May 5, 2025

Abstract

Lucas (1986) viewed the (in)stability of expectations as an experimentally testable hypothesis. We use learning-to-forecast experiments (LtFEs) to study the stability of expectations in the lab in a multivariate setting. The (in)stability in the lab is characterized by the leading eigenvalues of the expectations feedback system. This empirical characterization, however, is different from what most learning theories would predict. Using a behavioral mixed heuristics switching model (HSM) calibrated to our LtFE we find an instability region inside the unit circle, with real or complex near unit root eigenvalues. Such experimental economies are unstable with subjects in the lab coordinating on trend-following behavior amplifying temporary bubbles and crashes.

1 Introduction

The stability of equilibria in an economy depends on the way agents form expectations. Lucas (1986) pointed out that stability is not just a matter of a theory of adaptive learning, but should be viewed as an *experimentally testable hypothesis*: *"I think an appropriate stability theory can be useful in weeding out these uninteresting equilibria ... But to be useful, stability theory must be more than simply a fancy way of saying that one does not want to think about certain equilibria. I prefer to view it as an experimentally testable hypothesis, as a special instance of the adaptive laws that we believe govern all human behavior."* (Lucas, 1986, pp. S424 - 425).

In response to Lucas' view in a series of papers Marimon and Sunder (1993), Marimon et al. (1993), Marimon and Sunder (1994) and Marimon and Sunder (1995) introduced *Learning-to-Forecast Experiments* (LtFEs) to study expectations and learning in a controlled laboratory setting. They used an overlapping generations economy, but since then LtFEs have been designed to study the (in)stability of different experimental economies with expectational feedback, such as asset pricing, cobweb and the New Keyensian framework. Hommes (2021) provides a recent overview[1].

A surprising finding in LtFEs is that (in)stability in the lab generally does not coincide with (in)stability under learning theory. Asset markets, for example, which are stable under many learning algorithms, are found to be unstable in the lab with asset prices following bubbles and crashes (Hommes et al., 2005). Coordination on bubbles even arises in large groups of up to 100 subjects (Hommes et al., 2021).

The strength of the feedback from expectations to realisations is an important characteristic of aggregate market behaviour (Sonnemans and Tuinstra, 2010). In particular,

---

[1]Recent LtFEs in NK macro models include Adam (2007), Arifovic and Petersen (2017), Pfajfar and Žakelj (2018), Hommes et al. (2019a), Hommes et al. (2019b), Gali et al. (2021), Kryvtsov and Petersen (2021) and Assenza et al. (2021) . LtFEs are in fact repeated Keynes' beauty contest or guessing games introduced in Nagel (1995). See Mauersberger and Nagel (2018) for a survey on beauty contest games in the lab.

Bao and Hommes (2019) show that the (in)stability of housing market LtFEs varies with the discount factor, which coincides with the eigenvalue of this simple 1-dimensional expectations feedback system. For near unit root eigenvalues (e.g., 0.95) the housing market is unstable and exhibits large bubbles and crashes, while for smaller eigenvalues (e.g., 0.7) the market is stable and subjects in the lab quickly learn the rational expectations fundamental price.

The economy, however, is a higher dimensional system. The expectations people have of one variable tend to depend on various other variables, and the realisation of a variable depends on expectations of various others. How do subjects in the lab learn and what determines (in)stability in a higher dimensional economy?

A simple example of a 2-dimensional macroeconomic framework is the basic New Keynesian (NK) model, where agents form expectations about inflation and output gap (Woodford, 2003; Galí, 2008). Assenza et al. (2021) designed a LtFE for the NK model with different treatments depending on the strength of the inflation targeting monetary policy rule. They show that the NK experimental economy is unstable when the inflation targeting rule is weak and inflationary spirals may arise even when the Taylor principle holds (i.e., the interest rate responds more than 1-1 to inflation). Only when the inflation targeting coefficient is sufficiently strong (e.g., 1.5) the economy becomes stable in the lab, with inflation and output gap fluctuating close to their target and RE equilibriium.

But what is the mechanism driving these results? In this paper we show that the (in)stability of experimental economies can be characterized by the eigenvalues of the expectations feedback system, but that the characterization is different from what most learning theories predict. To better understand their role, we systematically vary the eigenvalues in a LtFE in a simple abstract expectations feedback setting. We focus on two-dimensional systems, such as the NK framework, but our results can be applied to higher dimensional economies by considering the 2-dimensional sub-system using the two leading eigenvalues (i.e., largest in absolute value).
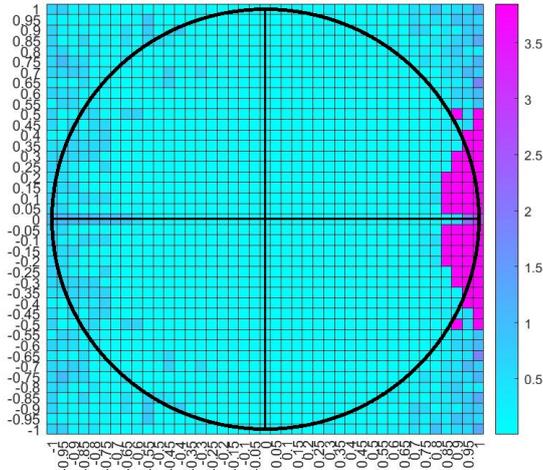
Figure 1: Heatmap showing the relative absolute distance (RAD) simulations using the mixed heuristics switching model (HSM) for a grid of eigenvalues in the complex unit square. The x-axis shows the real part, the y-axis the complex part of the (leading) eigenvalues. The unit circle is shown for clarity, illustrating the stability region of typical learning theories. The purple region is our predicted *instability region* in the lab.

The *heatmap* in Figure 1 provides an empirical answer to Lucas' stability question, with the blue color representing stable behavior in the lab and the purple color unstable behavior. Learning theory typically predicts stability when the eigenvalues of the expectations feedback system are inside the unit circle. However, this contradicts with the purple instability region in the lab, which lies (partly) inside the unit circle. The heatmap is obtained from simulations of the relative absolute deviation (RAD) from the fundamental value, a stability measure introduced by Stöckl et al. (2010), using a behavioral mixed heuristics switching model (HSM), generalizing the HSM in Anufriev and Hommes (2012). The mixed HSM is calibrated using data from our LtFE with different treatments corresponding to a small, carefully chosen set of eigenvalues; see Section 2 for details. The heatmap shows that the region of instability within the unit square is roughly given by a rectangular region of (complex) eigenvalues λ, for which $0.85 \leq \text{Re}\{\lambda\} \leq 1$ and $-0.4 \leq \text{Im}\{\lambda\} \leq 0.4$. The instability region includes e.g. the real near unit root eigenvalue 0.95, consistent with the bubbles in the asset pricing LtFEs in Hommes et al. (2005). On

the other hand, in the NK framework an aggressive inflation targeting Taylor rule (with coefficient 1.5) corresponds to complex eigenvalues $\text{Re}\{\lambda\} \approx 0.79$ and $\text{Im}\{\lambda\} \approx 0.24$, which lies very close to the stable region, consistent with the LtFEs of the NK framework in Assenza et al. (2021).

The paper perhaps closest to ours in the literature is Anufriev et al. (2022), who study the role of eigenvalues in the context of a two-dimensional-repeated guessing game. However, there are a number of important differences. Firstly, their subjects must forecast 1-period ahead, corresponding to the timing of expectations in a cobweb economy. In contrast, in our experiments we use a temporary equilibrium set-up, as in asset pricing or NK models, where predictions are two steps ahead. This creates the possibility for rational bubble solutions, as discussed in Hommes et al. (2005). Secondly, Anufriev et al. (2022) consider four treatments all with real eigenvalues (a sink, a source or a saddle point, with positive or negative leading eigenvalue), they do not consider the important case of complex eigenvalues arising e.g. in the NK model. Finally, Anufriev et al. (2022) provide participants with perfect information about the underlying laws of motion, while our subjects only have qualitative information about the market, mimicking real world conditions. For example, our subjects only know the signs (positive or negative) of the expectations feedback system, not their magnitude.

The paper is structured as follows. Section 2 presents the design of the experiment. Section 3 compares the observed aggregate market behaviour in the different treatments. In Section 4 we investigate which learning models best replicate the decisions the participants made in the lab. Section 5 develops a new mixed HSM in order to be able to predict the qualitative market dynamics ex ante. We lay out the evidence for using eigenvalues rather than only positive/negative feedback as predictors in Section 6. Finally, Section 7 concludes.

2 Experimental set-up

In the experiment individuals made decisions privately on a computer. They were randomly divided into groups of 6 and repeatedly submitted two numbers: a prediction for the price of some good x and a prediction for the price of some good y. The expectations of all members in the group determined what the next prices were. The participants did not know exactly how the prices were calculated, they only had qualitative descriptions of the relationships between predictions and realisations, namely whether the expectations feedback is positive or negative. An example of a chosen model, with eigenvalues $(0.95, 0.5)$, is:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.725 & 0.225 \\ 0.225 & 0.725 \end{pmatrix} \begin{pmatrix} x_{t+1}^e \\ y_{t+1}^e \end{pmatrix} + \begin{pmatrix} 0.65 \\ 0.65 \end{pmatrix}.$$

Once the prices had been determined they were revealed to all participants before they made their next predictions. This process was repeated for fifty time-steps. The experimental instructions presented to the subjects can be found in Appendix A.

Predictions were always made for time $t+1$, before the prices at time t were revealed. Subjects had all prices up till time $t-1$ available to them when forming their expectations. The first two sets of predictions therefore had to be made without any observed prices. To give some guidance, these initial predictions all had to be between 0 and 20. After the predictions of the first two time steps were made, the first prices were revealed. From then onwards the participants were only told that their predictions had to be non-negative. If, however, they attempted to submit predictions higher than 100, they were informed of an upper bound of 100.

Participants were rewarded based on the accuracy of their predictions. Earnings depended on both variables in every round equally. We determined the prediction error in each variable, added them up to find the total prediction error, and then use a hyperbolic

pay-off function similar to Hommes et al. (2019a):

$$\frac{100}{1 + 0.5(|x - x_e| + |y - y_e|)}.$$

One 'point' from the hyperbolic pay-off function was worth 1 eurocents. To ensure no one was under-compensated for their time (approximately 2 hours) we added a participation fee of €10.

To set the incentives even higher we build on the literature on loss aversion, and we framed the pay-off as a loss rather than a gain (Kahneman and Tversky, 1984) . Participants were informed at the beginning of the experiment of the maximal earnings and were told that these earnings would shrink when prediction errors were made. The pay-off function is not affected by this framing but the instructions are different from those in other similar experiments. See Appendix A.3 for this part of the instructions.

We chose the expectations feedback system that determines the prices based on the predictions to be linear. Understanding first a linear model is helpful since non-linear models will usually admit a linear approximation around a steady state. We also chose a model without noise since a two-dimensional setting could already be expected to be hard to understand even in the deterministic limit. We aimed for a setting that was as simple as possible.

2.1 Treatments

We used five different treatments to study the effect of eigenvalues of the expectations feedback model on market stability. The treatments differ in the linear functions from predictions to prices. These are the functions f which determine the relationship (see Appendix D for the linear maps of all treatments): :

$$(x_t, y_t) = f(\bar{x}^e_{t+1}, \bar{y}^e_{t+1}).$$

Note that the realized prices x_t and y_t both depend on the average forecasts $\bar{x}^e_{t+1}$ and $\bar{y}^e_{t+1}$ of all individuals in the group.

We looked only at eigenvalues with positive real part[2]. For the real pairs, we kept the non-leading eigenvalue constant at 0.5. For the leading eigenvalue we then have chosen 0.7, 0.95 and 1.1. With the first two we aim to discover whether we see a change from stable to unstable behaviour before leaving the unit circle, as we do in the one dimensional case. The final eigenvalue allows us to test the new possibility of a saddle point. For the complex eigenvalues we want to be able to make a comparison with dynamics under real eigenvalues with the same absolute value. We look at absolute values 0.7 and 0.95, both at angle $\frac{\pi}{4}$. We will refer to the three treatments with real eigenvalues as $0.7real$, $0.95real$, $1.1real$ and the two treatments with complex eigenvalues as $0.7complex$ and $0.95complex$ respectively. This gives us five treatments, as illustrated in Figure 2.

We chose the fixed point to be at $(13, 13)$ for all treatments. For the treatments with real eigenvalues we keep the eigenvectors constant as in Figure 3. By fixing this we have no further free parameters in the model and impose a consistent symmetry for simplicity. The dotted line, $x = y$, indicates the direction corresponding to the leading eigenvalue (so the solid line indicates the eigenvector corresponding to 0.5 every time). For the complex eigenvalues it is not possible to have these eigenvectors. We chose the models for the complex treatments to have one negative entry in the top-right position, since this setting arises in the linearised New-Keynesian model due to the negative feedback from the interest rate on the output gap (see Assenza et al., 2021).

Finally we performed a sixth treatment as a robustness check, $0.95robust$, where the eigenvalues are the same as in $0.95real$ but the model is different. This will be discussed in Section 6.

The models for all treatments can be found in Appendix D.

---

[2]This case appears in the NK model. Furthermore, Heemeijer et al. (2009) and Anufriev et al. (2022) show that negative real eigenvalues lead to stable behavior.
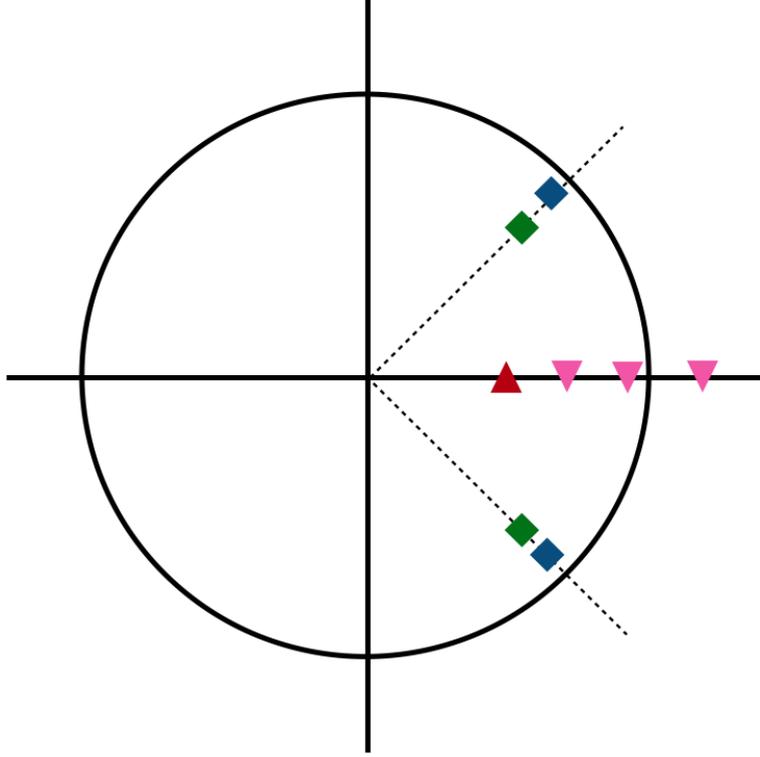
Figure 2: Eigenvalues in the five treatments. The red triangle at 0.5 pairs with the various pink ones (0.7, 0.95, 1.1) to form the real pairs, while the two green and two blue squares represent the complex pairs with the same absolute values of 0.7 and 0.95.

2.2 Limits

We limit predictions to $[0, 100]$, but for the $1.1real$ treatment it is possible that the realised price falls outside that interval despite predictions being in it. To prevent frustration with participants we do not permit this, and change any price below 0 into 0 and above 100 into 100. In other words, given the functions f in Appendix D, the price is actually given by:

$$(x_t, y_t) = \min\left(\max\left(f(\bar{x}^e_{t+1}, \bar{y}^e_{t+1}), 0\right), 100\right)$$

This means that for the $1.1real$ treatment we create two extra steady state equilibria: $(0,0)$ and $(100, 100)$. These could be seen as proxies for divergence to $-\infty$ and ∞ respectively.
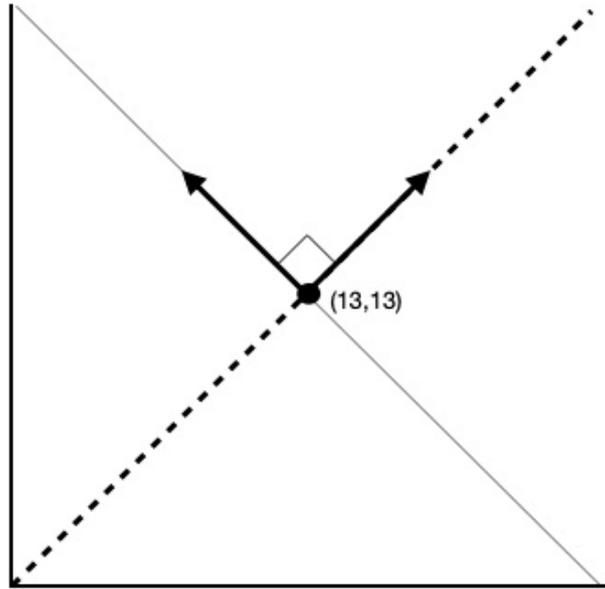
Figure 3: Eigenvectors kept fixed for all real treatments. The dotted line, $x = y$, indicates the direction corresponding to the leading eigenvalue. The solid line indicates the direction corresponding to eigenvalue 0.5.

2.3 Implementation

The experiment was run in the CREED laboratory at the University of Amsterdam in April 2018 and April 2019. For the treatments $0.7real$, $0.95real$ and $1.1real$ we had six groups with six subjects each, and for the treatments $0.7complex$, $0.95complex$ and $0.95robust$ we had seven groups with six subjects each. This brought the total number of subjects to 234. No subject participated in more than one session. A session lasted about 2 hours in total. Earnings, including the €10 participation fee, averaged €33.24.

The experiment was programmed in oTree (Chen et al., 2016). Before the experiment began, participants read instructions on screen and answered a series of control questions to ensure they understood what they had read (see Appendix A for instructions).

3 Aggregate dynamics

Figure 4 shows the market prices of x and y side by side, for all groups, sorted by treatment. A first, perhaps surprising result is that, despite limited information, a number of

treatments is stable with subjects able to learn to coordinate on the steady state equilibrium. We see clear differences between the treatments. $0.7real$ results in more stable markets than $0.95real$ and $1.1real$. In the case of $0.95real$ we observe rather persistent oscillations in the various groups, albeit of different amplitudes, in line with the earlier 1-dimensional LtFEs. In treatment $1.1real$ the groups showed quite different dynamics. Some groups coordinated on the $(100, 100)$ equilibrium, others on $(0, 0)$, which can be seen as a proxy for divergence to $+\infty$ and $-\infty$ respectively. Others had oscillatory dynamics.[3] Notably, $0.7complex$ and especially $0.95complex$ seem to have more stable markets than their real counterparts.

Another way to look at the observed prices is by drawing phase plots. Figure 5 shows a typical example of a phase plot for each treatment. We see what we would expect based on the eigenvalues and vectors: for the real treatments the fixed point is approached in a straight line following the $x = y$ direction (the weakly stable/unstable manifold), while in the complex treatment the fixed point is approached along a spiral. This result for the real treatment can be seen as a result of our choice for the manifolds, but it can also be that $x = y$ is simply a focal point. We will discuss this further in later sections, but it is important to note this for the further analysis where we see a much bigger difference between measures for x and for y for the complex treatments than for the real treatments. We also emphasise this result for the saddle point treatment $1.1real$, where the dynamics followed the unstable rather than the stable manifold, probably contributing to the lack of convergence to the interior equilibrium.

---

[3]Looking at the different results for the unstable eigenvalue 1.1 these seem to be driven by individual participants making idiosyncratic predictions which threw off the rest of the group. In the group where things seemed to have stabilised at 100, one participant suddenly decided to put a very low prediction (potentially out of boredom) resulting in the price dropping and all participants changing their predictions. In the group where we see something close to convergence there is actually 1 participants putting huge fluctuations in (switching between 0 and 60) and the other participants switching between high values to try and understand this behaviour. Yet another group includes a participant who simply started predicting one value (far from the equilibrium). It is likely that these 'rogue' participants were so common in this treatment due to its heavily unstable behaviour. Only one other group exhibited this type of erratic individual behaviour, see next footnote.
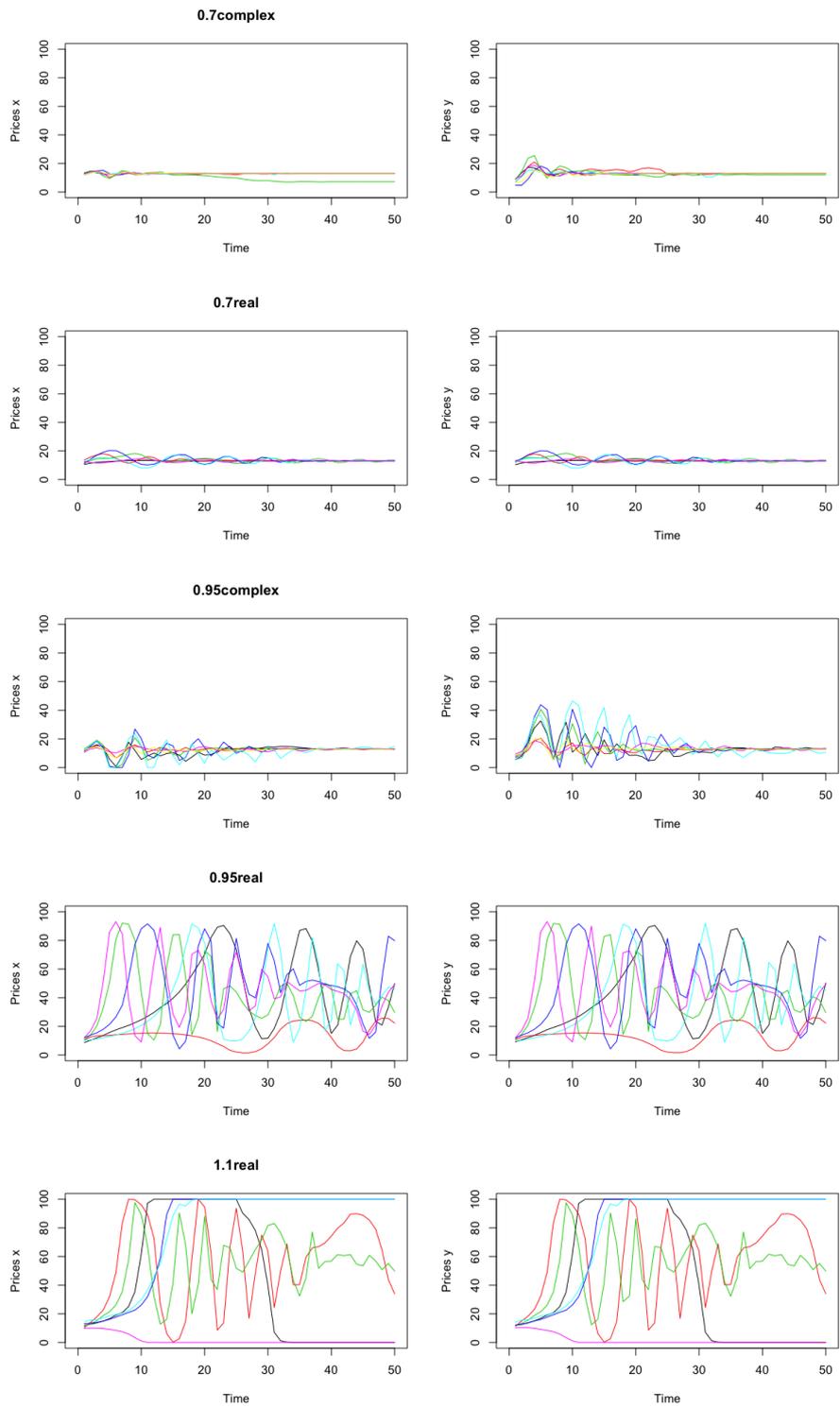
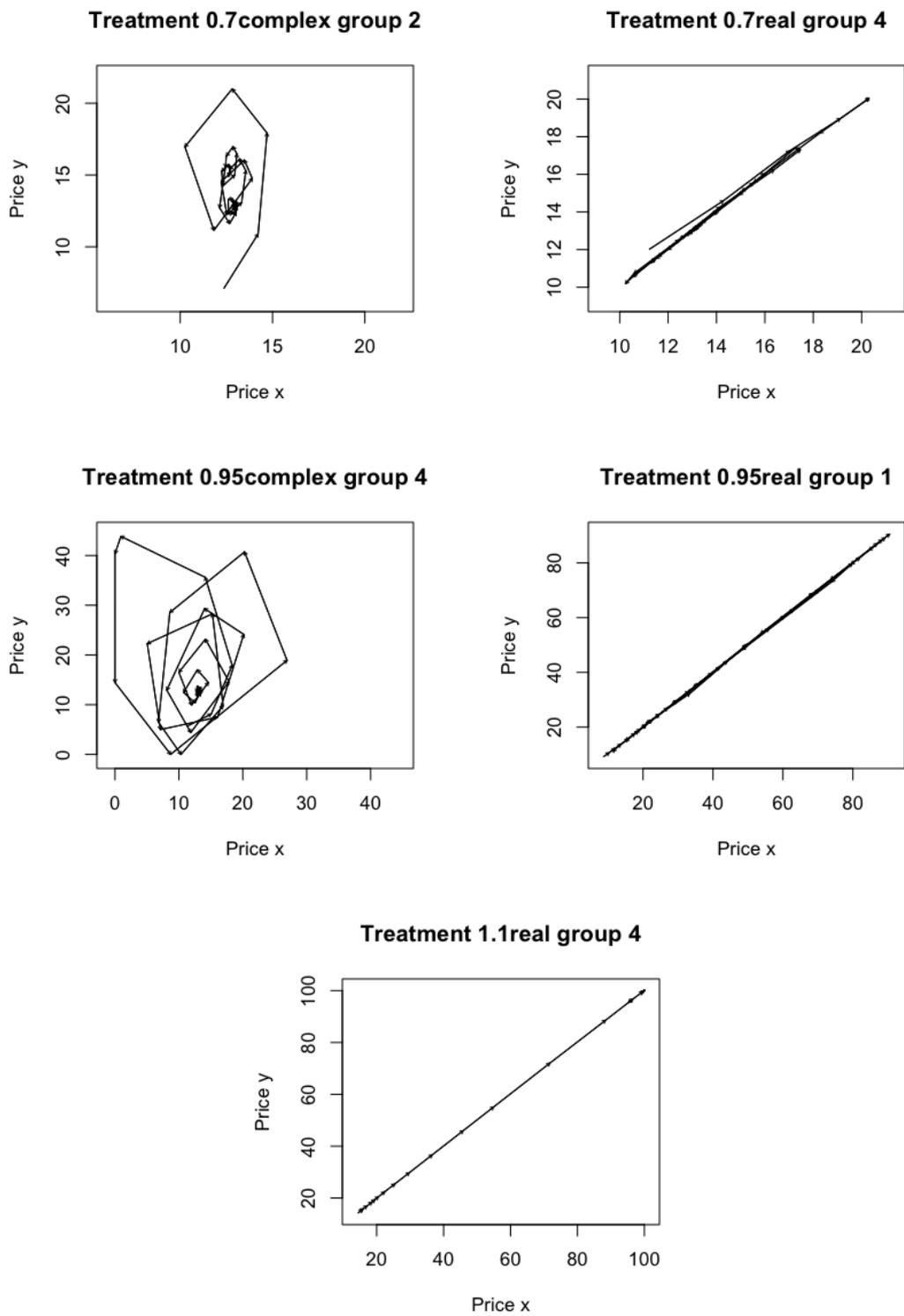Figure 4: Market prices of x (left) and y (right) in each group, sorted by treatment

13

Figure 5: Phase plot showing the observed prices for one exemplary group per treatment. Note that the scales may differ per treatment.

3.1 Measuring stability

In order to quantify the differences in stability we observe in Figure 4, we can use the bubble measures presented in Stöckl et al. (2010): the Relative Absolute Deviation (RAD) and Relative Deviation (RD) from the equilibrium $(13, 13)$. Furthermore, we also calculate the standard deviation of the prices, as that captures volatility in the saddle point treatment $1.1real$ as well. Accounting for a brief learning period, we calculate these measures from period 11 onwards.

$$RAD = \frac{1}{40} \sum_{t=11}^{50} \frac{|p_t - 13|}{13} \tag{1}$$

$$RD = \frac{1}{40} \sum_{t=11}^{50} \frac{p_t - 13}{13}, \tag{2}$$

$$std = \sqrt{\frac{1}{40} \sum_{t=11}^{50} (p_t - \bar{p})^2} \tag{3}$$

where p_t can be either x_t or y_t. We calculate the three measures for x and y separately and then take the average of those two.[4] Since there are three equilibria for the $1.1real$ treatment we do not apply the RAD and RD measures there. The averages per treatment can be found in Table 1.

Table 1: Average RAD, RD and std across groups per treatment

	RAD	RD	std
0.7complex	0.013	0.002	0.308
0.7real	0.054	0.013	0.992
0.95complex	0.138	−0.011	2.77
0.95real	2.186	2.084	19.200
1.1real	-	-	20.855

[4]$0.7complex$, group 3 has been removed for further analysis since there was one 'rogue' subject in this group. The participant switched to entering only 0 and 100 after a certain number of rounds, even when the price stabilised near the equilibrium. As a result, this participant made much less money than all the others in her group. The participant clearly did not try to maximise her earnings by attempting to predict the prices as well as possible.

We see that almost all treatments have a positive RD, which means that the price is on average above the equilibrium. For $0.7real$, $0.7complex$ and $0.95complex$ the RD is very close to 0 indicating stability of the fixed point. The RAD is also close to 0, but $0.95complex$ is larger here. The RAD and RD for $0.95real$ are much larger than for the other treatments, indicating its instability. The same pattern can be found for the standard deviation as for RAD, indicating $0.95real$ and $1.1real$ being more unstable than the other three treatments.

We use the Mann-Whitney rank sum test to see whether these differences and similarities are statistically significant (see Tables 2, 3).

Table 2: P values Mann-Whitney test comparing average Relative Absolute Deviation (RAD) and Relative Deviation (RD) in groups per treatment

	0.7complex	0.7real	0.95complex	0.95real
0.7complex		0.041	0.445	0.004
0.7real	0.132		0.101	0.009
0.95complex	0.002	0.138		0.005
0.95real	0.002	0.002	0.001	

Notes: The p-values below diagonal refer to the test results for RAD, above diagonal to the results for RD.

Table 3: P values Mann-Whitney test comparing average standard deviation (std) in groups per treatment

	0.7complex	0.7real	0.95complex	0.95real
0.7real	0.132			
0.95complex	0.001	0.073		
0.95real	0.002	0.002	0.001	
1.1real	0.065	0.065	0.051	0.937

We see that the difference between $0.95real$ and all other treatments is indeed statistically significant at 1% for all three measures, except for the difference in standard deviation from $1.1real$. Notably, this includes the difference between $0.95real$ and $0.95complex$. $0.95complex$ on the other hand is not significantly different from the $0.7real$ treatment by

either measurement on the 5%-level. We also see that $0.7real$ and $0.95real$ are different at a 1% level by all three measures showing a big shift in stability within the unit-circle. $0.7complex$ and $0.95complex$ are significantly different by the RAD and std measures, showing a shift in stability within the unit-circle for the complex values also when changing the absolute value. Considering the $1.1real$ treatment, we see that its standard deviation is weakly significantly different from those of treatments $0.7real$, $0.7complex$ and $0.95complex$, but not from $0.95real$. The reason that we only observe weakly significant differences for the first three treatments while the actual difference between the average standard deviation is substantial is that groups are more heterogeneous in behaviour in the $1.1real$ treatment (some show high oscillations, some show convergence to one of the boundaries, one group even switching between longer phases at the two boundaries).

3.2 Coordination

In order to get an idea of the level of coordination among participants we can look at the Dispersion Error. This is a measure of the heterogeneity in predictions among the participants in a group (Muth, 1961). Allowing for a brief period of learning, it is given by:

$$\frac{1}{40}\frac{1}{6}\sum_{t=11}^{50}\sum_{i=1}^{6}(p_{it}^e - \bar{p}_t^e)^2 \tag{4}$$

If this measure is high, there will be high heterogeneity in the predictions, and little coordination. Of course, we will see a very low dispersion error as soon as a price has converged. Table 4 shows the average dispersion error in each treatment. We see that indeed, $0.7real$ and $0.7complex$ have the lowest dispersion error, which can probably be explained by the stability in these treatments. $0.95complex$ on the other hand has a much higher dispersion error. This is somewhat surprising, but links to the findings in Heemeijer et al. (2009). They found that their more stable treatments could have high dispersion errors, and that in fact this lack of coordination could lead to stabilisation since there is

no coordination on trend-following heuristics and bubbles.

Table 4: Average dispersion error across groups per treatment

0.7complex	0.7real	0.95complex	0.95real	1.1real
0.577	0.222	12.616	30.080	12.948

Table 5: P-values Mann-Whitney test comparing average dispersion errors in groups per treatment

	0.7complex	0.7real	0.95complex	0.95real
0.7real	0.589			
0.95complex	0.051	0.181		
0.95real	0.041	0.041	0.234	
1.1real	0.041	0.065	0.345	0.818

We have again performed Mann-Whitney rank sum tests in order to assess whether the differences we see in the dispersion errors are statistically significant. The results are reported in Table 5. We see that the dispersion error of $0.95complex$ is not significantly different from that of $0.95real$ (0.234) and $1.1real$ (0.345), indicating again how little coordination we saw in this relatively stable treatment.

4   Models of expectation formation

We begin by trying to find a model that can replicate the predictions made by the participants in the experiment in each time step. This means that we give the model the whole set of information available to the participant at time t, and we see how close the prediction the model makes given this information is to the prediction the participant made at time t.

4.1   Adaptive rules

A first model we try to fit is some form of adaptive expectations. By this we mean a model of the form:

$$\begin{pmatrix} x_{t+1}^e \\ y_{t+1}^e \end{pmatrix} = \begin{pmatrix} x_t^e \\ y_t^e \end{pmatrix} + \mathbf{A}^T \begin{pmatrix} x_{t-1} - x_t^e \\ y_{t-1} - y_t^e \end{pmatrix} \tag{5}$$

where $\mathbf{A}$ is a $2x2$ matrix. We specify three cases that we will estimate:

- The 'one parameter model', where $A_{11} = A_{22}$ and $A_{12} = A_{21} = 0$

- The 'two parameter model', where $A_{12} = A_{21} = 0$

- The 'four parameter model', where none of the entries are restricted

We estimate this through OLS.

4.1.1 One parameter adaptive rule

We begin by fitting a one-parameter model to our data. We estimate the parameter separately for each treatment. The first columns of Table 6 shows the results. We see that the estimated parameters are smaller for the complex treatments than for the real ones. Note that when $A = 1$, the participants always predict simply the most recently observed price. We call this the naive model. When $A = 0$ the participants consistently enter their initial prediction. A lower A would thus correspond roughly to more constant predictions, while a higher A would correspond to more weight on recent observations. This is a first explanation for the lower estimated parameter in the complex treatments, since these were quite stable and thus had quite constant predictions. It does not, however, explain the high estimate for $0.7real$. This can be explained by the fact that $0.7real$ has lower frequency oscillations than the complex treatments do. As a result, recent observations are more informative and receive a higher weight. Note also that in the case when participants end up in equilibrium, both using the last observed price or their own last prediction would yield similar predictions, so any weights between the two could be justified in those instances.

Table 6: Estimation results for adaptive rules. Stars indicate significance at 0.001 level (***), 0.01 level (**) or 0.05 level (*).

Treatment	One Parameter	AIC (1)	Two Parameter	AIC (2)	Four Parameter	AIC (4)
0.7complex	0.58(***)	12132.168	0.534(***)	12124.667	0.497(***)	11586.309
			0.601(***)		0.115(***)	
					0.655(***)	
					-0.378(***)	
0.7real	0.731(***)	11917.063	0.725(***)	11918.847	0.712(***)	11910.162
			0.736(***)		0.014	
					0.615(***)	
					0.134(***)	
0.95complex	0.645(***)	25154.28	0.651(***)	25156.102	0.628(***)	24442.575
			0.643(***)		0.155(***)	
					0.664(***)	
					-0.326(***)	
0.95real	0.939(***)	27516.181	0.939(***)	27518.181	1.59(***)	27495.478
			0.939(***)		-0.65(***)	
					0.34(*)	
					0.602(***)	
1.1real	0.831(***)	27009.56	0.837(***)	27011.231	1.204(***)	26979.885
			0.825(***)		-0.376(***)	
					0.955(***)	
					-0.133(*)	

4.1.2 Two parameter adaptive rule

Columns 3 and 4 of Table 6 show the results for the two parameter estimation. The first thing to note is that the estimated parameters for x and y in the real treatments are almost equal, and that these are also almost equal to the estimated parameter in the one-parameter case. This is simply explained by the fact that the observed x and y were almost always equal in these treatments. For the complex treatments we observe slightly bigger changes in the parameters, but no clear trend emerges. The AIC does not change much for any of the treatments, indicating that the added parameter does not increase the explanatory power of the model much. We seem to be able to conclude that adding this second parameter does not offer significant gains. This indicates that participants may not have used different adaptive rules for predicting x and y.

4.1.3 Four parameter adaptive rule

Columns 5 and 6 of Table 6 finally show the results of an estimation with four parameters. For the real treatments we find some parameters with lower significance, and the AIC does not improve much. The added parameters again do not seem to add much explanatory power, which is unsurprising since x and y are generally equal for the real treatments. For the complex treatments we find that all parameters are significant at a 1% level, and that the AIC has dropped quite a lot. This model is thus an improvement to the one- and two-parameter models for the complex treatments. This indicates that for the complex treatments participants may have used information on x when predicting y and vice versa.

We will return to a comparison of the three types of adaptive rules in Subsection 4.3.

4.2 Heuristic Switching Model

A much more involved model of expectation formation, in which agents actually learn how to form their expectations as they go along, is the Heuristic Switching Model. This

was introduced in Brock and Hommes (1997) and found to successfully explain dynamics observed in one-dimensional LtFEs in Anufriev and Hommes (2012). We take the simplest way possible to generalise the one-dimensional model to two dimensions: we assume that agents only use information about x to predict x, and information about y to predict y. This is also the approach used in Assenza et al. (2021) for the 2-D NK model.

The idea behind the 1-D HSM is that there are four heuristics that participants are choosing from to make their predictions in each time step.[5] We will use the heuristics used in Anufriev and Hommes (2012), namely the Adaptive heuristic (ADA), Weak Trend-following Rule (WTR), Strong Trend-following Rule (STR) and the Anchoring and Adjustment rule (LAA):

$$p_{1,t+1}^e = 0.65 p_{t-1} + 0.35 p_{1,t}^e \tag{ADA}$$

$$p_{2,t+1}^e = p_{t-1} + 0.4(p_{t-1} - p_{t-2}) \tag{WTR}$$

$$p_{3,t+1}^e = p_{t-1} + 1.3(p_{t-1} - p_{t-2}) \tag{STR}$$

$$p_{4,t+1}^e = 0.5(p_{t-1}^{av} + p_{t-1}) + (p_{t-1} - p_{t-2}). \tag{LAA}$$

First, a measure of the success realised by each strategy in the last time step is calculated. We use the same measure that is used to determine the pay-off of a prediction. We then include some memory of previous success by the heuristic, governed by memory parameter $\eta = 0.7$. This brings the performance measure of heuristic i to:

$$U_{t-1,i} = \frac{100}{1 + |p_{t-1} - p_{t-1,i}^e|} + \eta U_{t-2,i}. \tag{6}$$

We then use a discrete choice model with asynchronous updating to determine what heuristics the participants will use in the next time step. The intensity of choice parameter

---

[5]These 4 rules represent four different types of behaviour frequently observed in 1-D LtFEs: monotonic convergence, oscillatory convergence, explosive behaviour and persistent oscillations.

β determines the extent to which participants will switch to the most successful rule, and the parameter δ determines how many participants are switching strategies in an individual time step. The fraction of participants using strategy i to make their next prediction is then equal to:

$$n_{t,i} = \delta n_{t-1,i} + (1 - \delta) \frac{e^{\beta U_{t-1,i}}}{\sum_j e^{\beta U_{t-1,j}}}. \tag{7}$$

We generally use the same form and parameters as in Anufriev and Hommes (2012), with one exception. For the Strong Trend Following Rule, which looks like:

$$p_{t+1}^e = p_{t-1} + \gamma(p_{t-1} - p_{t-2}), \tag{8}$$

we choose γ to be 1.8 rather than the usual 1.3. We found this to be a better fit both for the one step ahead forecasting we are considering here and for the fifty steps ahead learning that will be considered in Section 5.[6] We make the same choices as Anufriev and Hommes (2012) in initialising the model and start with 25% weight on each rule.

In our generalised 2-D model participants can use one heuristic for predicting x and a different one for predicting y at any time, but the parameters that determine the choice of heuristic for each variable will be the same. We will refer to this as the 'decoupled HSM'.

There are of course other ways to generalise the HSM to two dimensions. In particular, we could consider that participants use information about y when predicting x and vice versa, as in the four parameter adaptive model. We will be developing such a model in Subsection 5.3 and we will refer to it as the Mixed HSM. For the mixed HSM we use different initial conditions, making direct comparisons here slightly muddied. We will return to the reasons for this in Section 5.

---

[6]The original coefficient 1.3 was found in Hommes et al. (2005) in asset pricing experiments with stabilizing forces, such as a far from equilibrium stabilizing fundamental value robot traders. Without those fundamental robot traders the fitted trend following rules have higher coefficients.

4.3 Comparing 1-step ahead forecasting models

Now that we have laid out our various adaptive and heuristic switching models, we will compare a number of learning models based on their performance as measured by the mean squared error in Table 7. In addition to the models named before, we will be considering the 'rational' model, where participants constantly predict the equilibrium $(13, 13)$.[7] We will also consider the naive model, where participants always predict the most recently observed price. When separate estimations were done for x and y, we here report the average of the two MSEs.

A first observation from Table 7 is that the rational model performs relatively poorly for every treatment, including the stable ones. This is not too surprising as it has by far the least flexibility. The naive model however, which is still very simple, outperforms some of the more tailored models for some of the treatments. In particular, the naive model outperforms one or the other form of the HSM for the more stable markets, as well as the adaptive models for all treatments.

For the adaptive models we see that the change from one to two or to four parameters barely makes a difference for any of the treatments. This means that simply adding the flexibility of adaptive learning rules for x and for y does not help substantially explain the observed predictions.

The decoupled heuristic switching model performs relatively poorly for the more unstable treatments (but still better than the other rules), and relatively well for the most stable treatments. Considering the mixed HSM that will be laid out in Section 5 decreseas the performance on the stable treatments. We discuss the reasons for this in this later section as well.

---

[7]Note that in the case of the saddle point there are two other equilibria, but we here only consider the interior steady state.

Table 7: Mean Squared Errors of 1-step ahead forecasts for various models of expectation formation.

	0.7complex	0.7real	0.95complex	0.95real	1.1real
Rational	4.047	6.843	51.662	$1,440.892$	$2,999.755$
Naive	0.338	2.562	13.777	713.128	$1,165.270$
Ada, one par	3.760	8.220	82.664	$1,043.681$	$1,352.051$
Ada, two par	3.857	8.221	82.583	$1,043.680$	$1,351.258$
Ada, four par	4.328	8.874	77.858	$1,152.872$	$1,388.446$
HSM decoupled	0.107	0.237	9.511	125.635	103.194
HSM mixed	3.021	27.474	13.200	99.926	103.323

4.4 Individual estimations

Finally, we have also estimated learning models for individual participants rather than for a treatment pool as a whole.

We estimated models of the form:

$$p_t^e = c + \sum_{k=1}^{3} \alpha_k x_{t-k} + \sum_{l=1}^{3} \beta_l y_{t-l} + \sum_{i=0}^{2} a_i x_{t-i}^e + \sum_{j=0}^{2} b_j y_{t-j}^e \qquad (9)$$

which is a two dimensional version of the model estimated by Heemeijer et al. (2009). We used a model selection method called LASSO (Tibshirani, 1996) to determine what parameters are explanatory for individuals.

We found some patterns in the parameters that were deemed insignificant for the various participants and treatments. We look at the estimated models for the prediction of x. Labelling a parameter as unused when less than 6 people in a treatment used it in their predictions, we find that the third lag of the price of y is not used in any of the treatments except $0.95complex$ (in predicting x). The third lag of the price of x is not used in any of the treatments except $0.7complex$. The second lag on the prediction of both x and y are only used in the complex treatments, and the second lag of the price of y is only used in the complex treatments and $0.7real$. We conclude that more information is used to make predictions for the complex treatments than for the real treatments, both

more information from the past and more 'mixed' information about both variables. This matches our lower adaptive parameters, which showed that in the complex treatments there was less weight on more recent observations. It also matches the success of the four parameter model for these treatments.

5   Long run behaviour

If our 1-step forecasting models in Section 4 performed perfectly, we would be able to use them to predict the aggregate dynamics of a market ex ante, without feeding the model information about observations in the laboratory past the initial stage. However, since our 1-step ahead forecasting models still make mistakes, these mistakes can add up to reproduce qualitatively different aggregate dynamics in the medium to long run from what is observed. Since one of the goals of this type of research is to be able to predict the stability of equilibria for eigenvalues that have not been tested in the lab, we are interested in predicting aggregate dynamics given only the initial conditions of the experiment. In this section we therefore investigate how various models perform qualitatively when we let them make predictions for 50 time steps, given only the initial conditions from the lab, i.e. the realized prices from the first two periods as well as the average predictions of the market in the first period.

5.1   Successes and failures existing models

We will now consider to what extent the learning models we considered in Section 4 can qualitatively predict the aggregate market dynamics we observed.

5.1.1   Rational model

First of all, the Rational Expectations model would predict markets which are stable at $(13, 13)$. This may be a reasonably good description of the observed dynamics (albeit

not the path of convergence) for $0.7real$ and $0.7complex$, and perhaps even the long run behavior (after period 30) of $0.95complex$, but it can not explain what we observe for $0.95real$ and $1.1real$, nor why we see differences between treatments to begin with.

5.1.2 Naive model

When we consider the naive model, and 'plug' this into the laws of motion in Appendix D, what we end up with is exactly the expectations feedback model for which we chose our eigenvalues and eigenvectors. We will thus see a change from stable to unstable dynamics exactly when the absolute value of the leading eigenvalue exceeds 1, which contradicts the observed instability of the $0.95real$ treatment. The naive model also predicts that convergence takes exactly as long for the complex treatments as their real counterpart. This too contradicts our findings, most clearly when comparing $0.95real$ to $0.95complex$.

5.1.3 Adaptive model

When simulating the adaptive model for 50 time steps, we can only replicate the oscillating dynamics that we found in several treatments for the complex treatments. We do see that the adaptive model fully stabilises on an equilibrium within 50 steps for exactly the same treatments that we found to be stable in the lab. However, in the treatments where we found instability ($0.95real$, $1.1real$) the adaptive model predicts one peak and then a slow but monotonic convergence towards the equilibrium, which does not match our qualitative findings.

5.1.4 Decoupled HSM

As argued by Anufriev and Hommes (2012) for one dimension, we find that the HSM can reproduce the change in stability when moving from $0.7real$ to $0.95real$. This happens because the Strong Trend-following Rule performs very well in the case of $0.95real$, amplifying price oscillations and coordination on bubbles. For $0.7real$ there is less coordination

at the start, in particular no coordination on the STR, reducing bubbles. The HSM needs to be initialised with the fractions for each heuristic and the dynamics can change quite a lot when these are altered. It would be hard to defend to choose different initial fractions for different treatments. This makes it impossible to reproduce the observed dynamics in all treatments using the decoupled HSM. Generally, the HSM underestimates how stable the $0.95complex$ treatment is, and overestimates how unstable the $0.95real$ treatment is. The HSM can reproduce large oscillations in the $0.95real$ treatment by making the initial fraction of strong trend followers very large, but with such fractions the $0.95complex$ shows large persistent oscillations as well. To approximate the stability in the $0.95complex$ case we need to increase the fraction of the adaptive heuristic so much that the $0.95real$ treatment becomes very stable. Summarizing, the decoupled HSM can *not* simultaneously, i.e. with the same initial fractions, explain the unstable behavior in the $0.95real$ treatment and the stable behaviour in the $0.95complex$ treatment.

5.2 Evidence for a mixed model

The decoupled HSM came close to predicting all qualitative dynamics, but it underestimated the stability of $0.95complex$. We may be able to reproduce this stability in the HSM by including some mixing, so by allowing y to be used when predicting x and vice versa. A mixed HSM would be conistent with the estimated individual rules in Section 4.4.

5.2.1 Why mixing may stabilise

The reason why mixing may aid in reproducing the stability is as follows. The spiralling dynamic we have in the complex treatments means that x and y are only moving in the same direction roughly half of the time, see Figure 6. When the two are moving in opposite directions, taking both trend-movements into account may temper bubble extrapolation. When subjects only focus on one of the trends, the participant would simply extrapolate
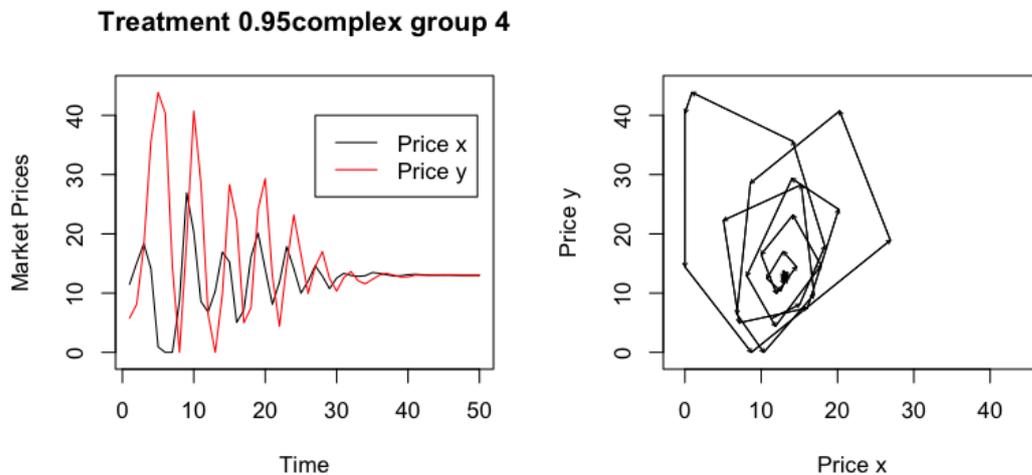
Treatment 0.95complex group 4

Figure 6: Example observed dynamics complex treatment 0.95*complex*.

that trend. In the case of two diverging trends, a form of mixed trend-following would take into account both the increasing trend of one and the decreasing trend of the other variable, and the overall result would be weaker trend-extrapolation.

For the real treatments, since here we found x and y to be equal almost all the time, mixing will not make a difference as long as the sum of the weights on x and y does not change.

5.2.2 Evidence mixing

We found evidence of the participants mixing the history of x and y for complex treatments in Section 4. First of all, the four parameter adaptive model performed better than the two parameter model for the complex treatments. The difference between these two was the inclusion of mixing. We also find that in the four parameter adaptive model for the complex treatments, all variables have small p-values, indicating that allowing for four parameters contributes to the predictive power of the model.

In the individual estimations we find further evidence of mixing. For the complex treatments we see that more information on both variables, including the other lagged

variables, is used to make predictions than for the real treatments.

5.3   Proposed mixed HSM

One may argue that there are many ways in which we may adapt the HSM to include
mixing. We choose the simplest adaptation of the decoupled HSM that would replicate
the qualitative dynamics in all treatments.

It turns out that keeping anchors as they are, and splitting adjustments equally across
the two variables fits the data well. This means that when predicting x, one looks at the
current price (or other anchor) of just x, but looks at the changes in both x and y to
forecast their adjustments. The four mixed rules are thus as follows:

$$p_{t+1}^e = 0.65p_{t-1} + 0.175x_t^e + 0.175y_t^e \qquad \text{(ADA-m)}$$

$$p_{t+1}^e = p_{t-1} + 0.2(x_{t-1} - x_{t-2}) + 0.2(y_{t-1} - y_{t-2}) \qquad \text{(WTR-m)}$$

$$p_{t+1}^e = p_{t-1} + 0.9(x_{t-1} - x_{t-2}) + 0.9(y_{t-1} - y_{t-2}) \qquad \text{(STR-m)}$$

$$p_{t+1}^e = 0.5(p_{t-1}^{av} + p_{t-1}) + 0.5(x_{t-1} - x_{t-2}) + 0.5(y_{t-1} - y_{t-2}) \qquad \text{(LAA-m)}$$

Now if we combine this with the right initial conditions, we can replicate the observed
dynamics. For the initial conditions, we are mainly constrained by the large instability
in the $0.95real$ case. The bigger we choose the initial fraction using STR, the better we
replicate this instability. But doing this does not give wrong qualitative dynamics for the
other treatments, as we saw occur in the decoupled HSM. We do however want to base
our initial fractions on our observations from the lab. Since we found that about 70% of
initital predictions either increased or decreased from the first to the second time step,
despite the fact that no new information came through, we conclude that there was some
ex ante strong trend extrapolation present. We use this as a proxy and thus place an
initial weight of 0.7 on the STR and split the further weight equally across the three other

heuristics.

Figure 7 shows for one exemplary group per treatment how the realised and simulated prices compare. These are representative of the fit we see in the other groups, except for 1.1*real*. For 1.1*real* the mixed HSM always predicts that the market will shoot to the $(100, 100)$ equilibrium while in some of our treatments we saw different behaviour, although there is agreement on the lack of convergence to the internal equilibrium.

5.4   Prevailing rules

Table 8 shows for the mixed HSM what fraction of agents uses what rule on average over time, over the two prices and over each market, per treatment. We see that the ADA rule prevails in the complex treatments and is used much less often in the real ones. This is in line with the success of the four parameter adaptive model we saw in the one step ahead estimation. Since the ADA rule is stabilising this helps explain the stability observed in the complex treatments. In the real treatments the STR is common, and here we also see the LAA doing quite well. Since the LAA includes trend following as well as a strong anchoring, it fits both the unstable dynamics of 0.95*real* and the stable ones of 0.7*real*.

Table 8: Average fraction using each rule for fifty step ahead mixed HSM

	0.7complex	0.7real	0.95complex	0.95real	1.1real
ADA	0.636	0.086	0.627	0.045	0.145
WTR	0.117	0.152	0.111	0.040	0.229
STR	0.211	0.283	0.208	0.651	0.585
LAA	0.036	0.480	0.054	0.264	0.042

5.5   Predicting (in)stability in the lab

Now that we have found a behavioral model for expectation formation that can qualitatively replicate almost all of the behaviour we observed in the lab, we can use this model
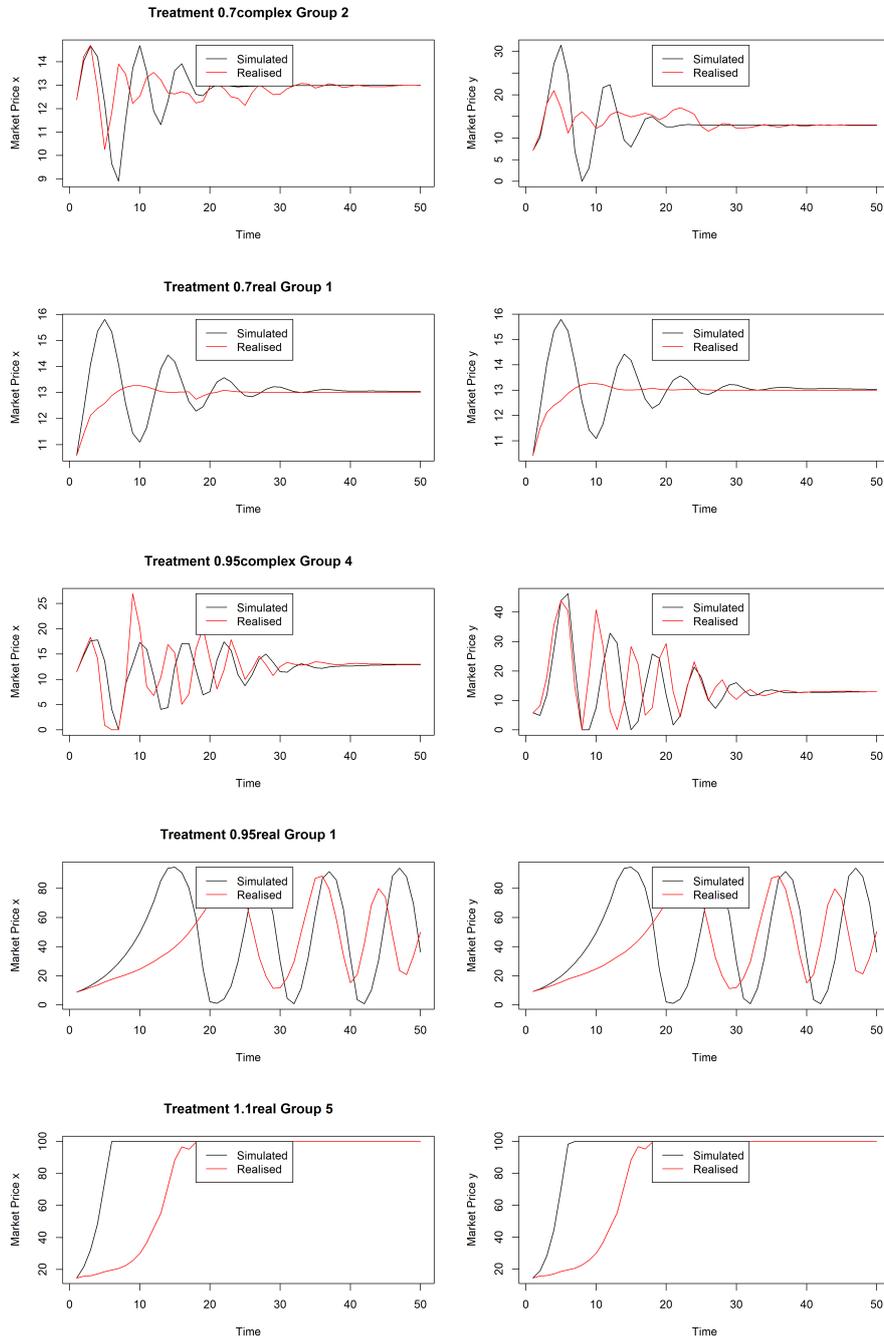
Figure 7: Realised prices and prices as simulated using the mixed HSM, 50 steps ahead, for both x (left) and y (right) for one exemplary group per treatment. Initial weight of 0.7 on STR, and 0.1 on every other heuristic.

to predict stability of settings we have not studied in the lab. We have simulated the behavioral model for pairs of eigenvalues across the complex unit square. For the real eigenvalues we use the same eigenvectors (x=y and x=-y), equilibrium $(13, 13)$ and stable eigenvalue $(\lambda = 0.5)$ as we did for our real treatments.

For complex eigenvalues we use the same form that we also used in the experiment for the complex treatments:[8]

$$
\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_{t+1}^e \\ y_{t+1}^e \end{pmatrix} + \begin{pmatrix} 10 \\ -10 \end{pmatrix}.
\tag{10}
$$

To calculate the parameters a, b, c, d we use the two characteristic equations together with the steady state equations for $(13, 13)$.

For each pair of eigenvalues we calculate the RAD as a measure of (in)stability. For each eigenvalue pair we ran 100 independent simulations where for each simulation we sampled six participants initial predictions randomly from the observations in the lab. We averaged those observations, as the experiment also takes averages of the six participants' predictions. Furthermore, we left the first 25 periods out of the calculation of the RAD to minimise the influence of the initial conditions.[9] The heatmap in Figure 8 in the introduction shows the results.

The blue region corresponds to stable, the purple region to unstable behavior. Overall, we find a roughly rectangular region of instability with (complex) eigenvalues λ, $0.85 \leq \mathrm{Re}\{\lambda\} \leq 1$ and $-0.4 \leq \mathrm{Im}\{\lambda\} \leq 0.4$. This instability region lies (partly) inside the unit circle, contradicting learning theories, e.g., naive, adaptive and sample mean expectations. Our predicted heatmap is based on the mixed HSM calibrated to the five treatments of our LtFEs, with eigenvalues given in Figure 2. The next section investigates how well our

---

[8]Rather than forcing the upper right entry to be negative for the complex eigenvalue matrices, we fixed the constant in the equations to be equal to the constant vector $(10, -10)$ we chose for our initial complex treatments.

[9]Corresponding to the learning phase applied in Section 3, we ran robustness checks with simulations leaving out only the first 10 periods. Results hardly change.

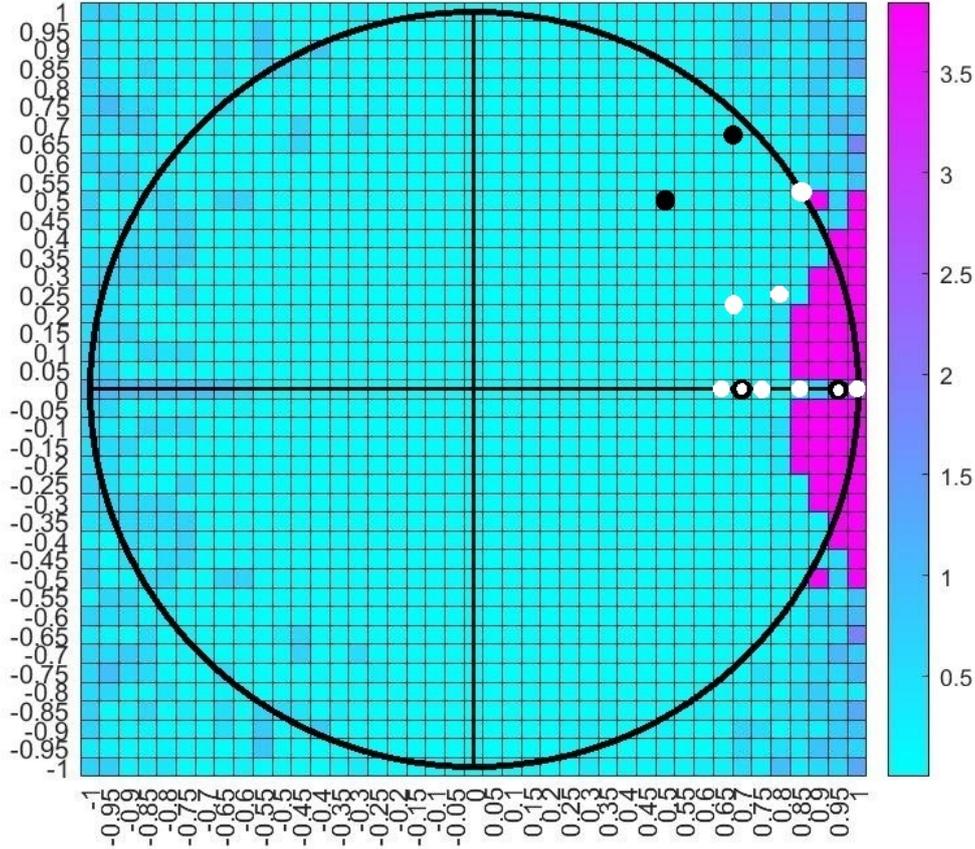predicted heatmap matches other LtFEs in the literature.

Figure 8: Heatmap showing the RAD simulated by the mixed HSM for a grid of eigenvalues in the complex unit square. The x-axis shows the real part, the y-axis the complex part of the (leading) eigenvalues. The unit circle is shown for clarity. The black dots are those from our experiment, the other dots are from LtFEs in Bao and Hommes (2019), Assenza et al. (2021), Hommes et al. (2019b), Hommes et al. (2019a), Hommes and Makarewicz (2021) and Anufriev et al. (2022); see also Table 9.

5.6 Behavior in other LtFEs

Figure 8 shows our heatmap together with the (leading) eigenvalues from the LtFEs literature. We include 1-, 2- and 3-D systems, linear and nonlinear, buffeted with small IID shocks. The relevant papers and corresponding eigenvalues are summarized in Table 9

Table 9: Observed lab behavior in other LtFEs

Reference	Model	Treatments	Leading eigenvalues	Observed lab behavior
Bao and Hommes (2019)	housing (1D linear)	high supply elasticity	$\lambda = 0.7$	stable ✓
		low supply elasticity	$\lambda = 0.85$	unstable, oscillations ✗
		no supply elasticity	$\lambda = 0.95$	unstable, large bubbles ✓
Assenza et al. (2021)	linear NK (2-D)	T1: $\phi_\pi = 1$	$\lambda_1 = 1, \lambda_2 = 0.76$	unstable, exploding ✓
		T2: $\phi_\pi = 1.005$	$\lambda_1 = 0.99, \lambda_2 = 0.77$	unstable, exploding ✓
		T3: $\phi_\pi = 1.015$	$\lambda_1 = 0.98, \lambda_2 = 0.83$	unstable, oscillations ✓
		T4: $\phi_\pi = 1.5$	$\lambda_{1,2} = 0.79 \pm 0.243i$	stable ✓
Hommes, Massaro & Weber (2019)	linear NK (2-D)	inflation targeting ($\bar{\pi} = 3.5$)	$\lambda_{1,2} = 0.79 \pm 0.243i$	stable, oscillations ✓
		inflation + output targeting	$\lambda_1 = 0.76, \lambda_2 = 0.66$	stable, oscillations ✓
Hommes, Massaro, & Salle (2019)	nonlinear NK (2-D)	target steady state	$\lambda_{1,2} = 0.67 \pm 0.22i$	locally stable ✓
		low inflation steady state	$\lambda_1 = 1.52, \lambda_2 = 0.65$	locally unstable, exploding deflationary spirals ✓ ?
Hommes & Makarewicz (2021)	nonlinear NK with PLT (3D)	inflation targeting	$\lambda_1 = 0.886, \lambda_2 = 0.287$	stable or small oscillations ✓
		weak PLT	$\lambda_{1,2} = 1.077 \pm 0.202i, \lambda_3 = 0.21$	unstable, oscillations ✓ ?
		strong PLT	$\lambda_{1,2} = 0.851 \pm 0.524i \ (|\lambda_{1,2}| = 0.99), \lambda_3 = -0.483$	stable ✓
Anufriev, Duffy & Panchenko (2022)	cobweb type (2-D)	sink	$\lambda_1 = 0.67, \lambda_2 = -0.5$	stable ✓
		positive saddle	$\lambda_1 = 1.5, \lambda_2 = -0.67$	unstable ✓ ?

Notes: The last column contains observed lab behavior in the given paper followed by a sign indicating whether the lab behavior is consistent with our predictions. ✓ stands for consistency, ✗ stands for inconsistency. For leading eigenvalues outside the unit square we add implied consistency together with *?* to indicate that we have not directly run comparable simulations.

and briefly discussed below. We find that our behavioral model correctly predicts the (in)stability in the lab in most of these LtFEs.

Bao and Hommes (2019) consider a linear 1-D housing model, where they vary the elasticity of housing supply. An increase in housing supply elasticity lowers the eigenvalue λ of the expectations feedback system. They consider three treatments:

1. $\lambda = 0.95$: large bubbles are observed in the lab;

2. $\lambda = 0.85$: no large bubbles, but oscillating prices around the fundamental value. Our model predicts stability rather than instability, although $\lambda = 0.85$ is near the instability region. This small difference may stem from two sources. First, our model is 2-dimensional, whereas the housing model is 1-dimensional. Second, in a 1-D system with eigenvalue $\lambda = 0.85$ participants might coordinate more easily on trend-following behavior in a single variable than in our 2-D simulations for eigenvalues $\lambda = 0.85$ and $\lambda = 0.5$;

3. $\lambda = 0.7$: stable behavior with prices converging to fundamental value.

Assenza et al. (2021) consider a 2-D linearized NK model with inflation and output gap expectations and an inflation targeting interest rate rule with target $\bar{\pi} = 2\%$ and coefficient ϕ_π. They consider four treatments depending on ϕ_π:

1. $\phi_\pi = 1$; $\lambda_1 = 1, \lambda_2 = 0.76$, with a continuum of steady state equilibria, and exploding paths for inflation and output;

2. $\phi_\pi = 1.005$; $\lambda_1 = 0.99, \lambda_2 = 0.77$, with the Taylor principle satisfied, a unique RE equilibrium, but exploding paths for inflation and output due to positive expectations feedback;

3. $\phi_\pi = 1.015$; $\lambda_1 = 0.98, \lambda_2 = 0.83$, with the Taylor principle satisfied and a unique RE equilibrium. This slightly more aggressive Taylor rule now adds negative feedback

from the interest rate to the output gap. This prevents exploding dynamics, but inflation and output gap still fluctuate around the RE target inflation and 0 output gap;

4. $\phi_\pi = 1.5$; $\lambda_{1,2} = 0.79 \pm 0.243i$ ($|\lambda_{1,2}| \approx 0.83$), with the Taylor principle satisfied and a unique RE equilibrium. This more aggressive Taylor rule adds sufficient negative feedback for the NK system to stabilize with inflation and output gap converging to the RE target inflation of 2% and 0 output gap. This stable behavior is consistent with our heatmap in Figure 8, as these complex eigenvalues are close to the boundary, but fall into the stable region.

In the same 2-D NK linear setting Hommes et al. (2019b) investigate whether targeting output adds to the stability of the NK economy. They ran a LtFE with two treatments:

1. $\lambda_{1,2} = 0.79 \pm 0.243i$: an interest rate rule that only targets inflation with coefficient $\phi_\pi = 1.5$. In most economies inflation and output gap fluctuate around their steady state values. This may seem surprising as this case has the same parameters as the stable treatment T4 in Assenza et al. (2021), except that the inflation target $\bar{\pi} = 3.5\%$, rather than the standard target $\bar{\pi} = 2\%$. A higher and non-standard inflation target may slow down convergence to the steady state. Notice further that this case is again close to the boundary between the stable and unstable regions, consistent with the somewhat mixed results.

2. $\lambda_1 = 0.76$, $\lambda_2 = 0.66$: inflation and output gap targeting, with coefficients $\phi_\pi = 1.5$ and $\phi_y = 0.5$. In the lab the behavior is more stable, with lower volatility of inflation and output gap in this treatment. Notice that this case is also close to the boundary and that the stable eigenvalue is 0.66, larger than our stable eigenvalue 0.5. Notice that in both treatments the eigenvalues are close to the boundary of unstable behavior. The additional output targeting moves the eigenvalues more into the stable region, hence these results are in line with ours.

Hommes et al. (2019a) ran LtFEs for a nonlinear 2-D NK model with two equilibria, the target steady state and a low inflation steady state. The target steady state has eigenvalues $\lambda_{1,2} = 0.67 \pm 0.22i$ and is locally stable in the lab. The low inflation steady state is a saddle with eigenvalues $\lambda_1 = 1.52$, $\lambda_2 = 0.65$; it is locally unstable in the lab with deflationary spirals emerging as soon as the path crosses the boundary (the stable manifold of the low inflation steady state) of the basin of attraction of the target steady state. This behavior in the lab thus seems to be in line with our (in)stability results.

Hommes and Makarewicz (2021) use LtFEs to compare inflation targeting (IT) versus price level targeting (PLT) in a nonlinear NK model. With inflation targeting the model is 2-D with target steady state eigenvalues $\lambda_1 = 0.886$, $\lambda_2 = 0.287$. This economy is stable in the lab with small oscillations in some cases, consistent with our results. Note that this eigenvalue is close to the boundary of the (in)stability region in the lab. In the case of PLT the model becomes 3-D. With a weak PLT rule, the leading eigenvalues are outside the unit circle, $\lambda_{1,2} = 1.077 \pm 0.202i$, $\lambda_3 = 0.21$ and the lab economy is very unstable with large fluctuations. With a strong PLT rule, the leading eigenvalues move inside the unit circle $\lambda_{1,2} = 0.851 \pm 0.524i$ ($|\lambda_{1,2}| = 0.99$) and the economy is stable in the lab, consistent with our results.

Finally, Anufriev et al. (2022) consider LTFEs as repeated beauty contest games. Although their timing of the expectations feedback is different from ours, they obtain similar results:[10]

1. A sink with eigenvalues $\lambda_1 = 0.67$ and $\lambda_2 = -0.5$ is very stable in the lab;

2. A saddle with positive eigenvalues $\lambda_1 = 1.5$, and $\lambda_2 = -0.67$ is very unstable in the lab.

---

[10]The timing in Anufriev et al. (2022) is $(x_t, y_t) = f(x_t^e, y_t^e)$, with 1-period ahead forecasts (as in the cobweb model), while our setup uses a temporary equilibrium framework $(x_t, y_t) = f(x_{t+1}^e, y_{t+1}^e)$, with 2-period ahead forecasts. Anufriev et al. (2022) also consider saddles with real negative eigenvalues, not considered here.

These results, for 1-D, 2-D and 3-D LtFEs, linear or nonlinear, are in line with our predicted (in)stabilty region as illustrated in the heatmap in Figure 8.

6 Robustness check: negative feedback

In the treatments discussed so far, we found a change from unstable to stable dynamics when we went from real eigenvalues to complex eigenvalues with the same absolute value 0.95. It is possible, however, that this result was driven by negative feedback, that is, the inclusion of a negative partial derivative. Equation (11) shows the model used for the treatment with real eigenvalues $(0.95, 0.5)$ and (12) the one for $0.95e^{\pm i\pi/4}$.

$$
\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.725 & 0.225 \\ 0.225 & 0.725 \end{pmatrix} \begin{pmatrix} x_{t+1}^e \\ y_{t+1}^e \end{pmatrix} + \begin{pmatrix} 0.65 \\ 0.65 \end{pmatrix} \tag{11}
$$

$$
\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.651 & -0.420 \\ 1.076 & 0.693 \end{pmatrix} \begin{pmatrix} x_{t+1}^e \\ y_{t+1}^e \end{pmatrix} + \begin{pmatrix} 10 \\ -10 \end{pmatrix} \tag{12}
$$

Assenza et al. (2021) ran 2D LtFEs in the New Keyensian model and argue that adding negative feedback through a more aggressive Taylor rule stabilises the NK system. One might therefore hypothesise that the signs of the partial derivatives rather than the eigenvalues should be used as predictors of stability. To test this hypothesis, we conducted an additional treatment where we had again the eigenvalues $(0.95, 0.5)$, but this time chose a model which included negative feedback, through one negative partial derivative, as in (12). If this would stabilise the dynamics it would suggest that eigenvalues are not sufficient as predictors, and we require the full picture of the various partial derivatives.

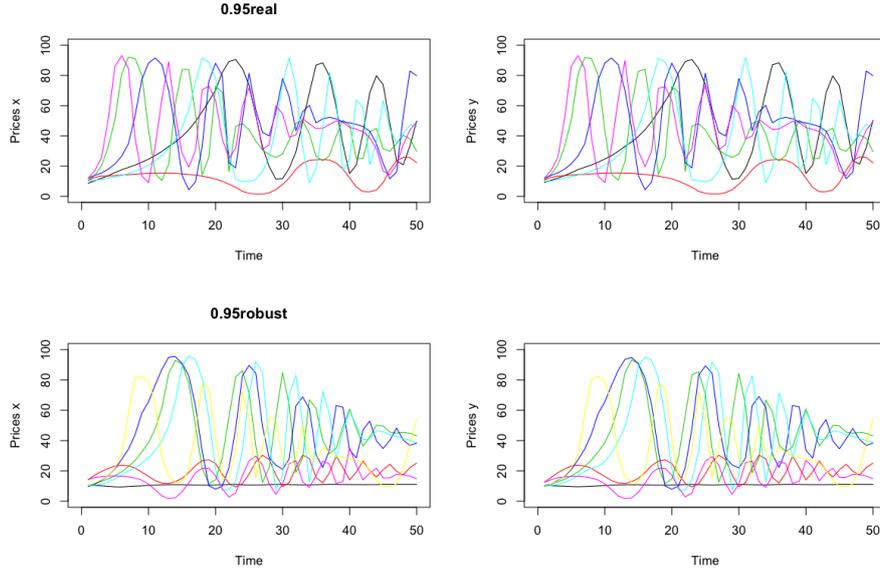The treatment $0.95robust$, with real eigenvalues 0.5 and 0.95, was thus conducted with

Figure 9: Market prices of x and y in each group, for the two treatments $0.95real$ (top panels) and $0.95robust$ (bottom panels).

expectations feedback model that includes negative feedback :

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1.15 & -0.2 \\ 0.65 & 0.3 \end{pmatrix} \begin{pmatrix} x_{t+1}^e \\ y_{t+1}^e \end{pmatrix} + \begin{pmatrix} 0.65 \\ 0.65 \end{pmatrix} \tag{13}$$

6.1   Results robustness check

Table 10: Bubble measures and average dispersion errors for 0.95real and 0.95robust

	0.95real	0.95robust
RAD	2.185	1.615
RD	2.084	1.495
std	19.200	15.065
Dispersion Error	30.080	15.011

Let us first focus on the observed prices, shown in Figure 9. We could roughly classify the observed markets in the $0.95real$ treatment as twice persistent large oscillations, once persistent small oscillations, and three times large oscillations which seem to lead to

convergence. In the $0.95robust$ we have four times large oscillations which seem to lead to convergence, twice small oscillations which seem to lead to convergence, and one very stable market. Based on this qualitative evidence one could say that $0.95robust$ appears somewhat more stable, but it is not entirely clear. In Table 10 we see that the RAD, RD and standard deviation for $0.95robust$ are indeed smaller than for $0.95real$. We perform a Mann-Whitney rank sum test to determine whether this difference is statistically significant. For all three measures the p-value of this test is 0.534, meaning that the difference is not statistically significant. The difference between the dispersion errors we observe is also not significant, with a p-value of 0.366. We thus conclude that the differences between the $0.95real$ and $0.95robust$ treatments are not significant. We find evidence that eigenvalues rather than individual partial derivatives are good predictors of stability.

7   Conclusion

We have run Learning to Forecast Experiments (LtFEs) in multivariate expectations feedback systems and characterized the (in)stability in the lab by its eigenvalues. Despite participants only having qualitative information about the underlying feedback system, we found that in multiple treatments the markets converged to the equilibrium value. In particular, $0.7real$, $0.7complex$ and $0.95complex$ converged to the steady state. This is especially notable since $0.95real$, which had the same absolute value as $0.95complex$, lead to very unstable markets. We find that complex eigenvalues with an angle of $\frac{\pi}{4}$ lead to more stable markets than real eigenvalues with the same absolute value.

The known models for 2D expectation formation we examined can not replicate this difference in stability between real and complex treatments. We introduce a version of the HSM where agents mix information on the two variables they are predicting when forming the adjustments to their anchors. This can explain stability for the complex treatments, where variables often move in opposite directions meaning the two trend-following adjust-

41

ments offset each other. We use this new behavioral model to make predictions for the stability of systems with eigenvalues that we did not test ourselves. We detect an instability region inside the unit circle, with real and complex near unit root eigenvalues, for which the economy in the lab is unstable, with subjects coordinating on trend-following behavior and bubbles and crashes. The stability and instability regions that we find are consistent with most of the other LtFEs in the literature.

We emphasized a simple, abstract and neutral formulation of the LtFEs underlying the stability region in the lab. In particular, our results apply to expectations feedback systems with small IID noise. We conclude with three important extensions for future work: First, how will framing affect the stability in the lab? For example, our subjects are asked to forecast two "prices" of an unspecified good. One can imagine that framing effects concerning what type of prices (e.g. stock prices, housing prices, consumer prices, inflation, etc.) may affect the stability in the lab. Second, the role of noise is important and it would be of interest to study e.g. how autocorrelated shocks may affect the stability in the lab. Finally, LtFEs can be unstable due to coordination on bubbles and crashes. It would be of interest to study how central bank communication may restore stability in the lab of an unstable feedback system. We leave these questions for future work.

References

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

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

Anufriev, M. and Hommes, C. (2012). Evolutionary selection of individual expectations

and aggregate outcomes in asset pricing experiments. *American Economic Journal: Microeconomics*, 4(4):35–64.

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

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

Bao, T. and Hommes, C. (2019). When speculators meet suppliers: Positive versus negative feedback in experimental housing markets. *Journal of Economic Dynamics & Control*, f107:103730.

Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica: Journal of the Econometric Society*, pages 1059–1095.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Galí, J. (2008). *Monetary policy, inflation, and the business cycle: An introduction to the New Keynesian framework*. Princeton University Press.

Gali, J., Giusti, G., and Noussair, C. (2021). Monetary policy and asset price bubbles: A laboratory experiment. *Journal of Economic Dynamics and Control*, 130:104184.

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

Hommes, C. (2021). Behavioral and experimental macroeconomics and policy analysis: A complex systems approach. *Journal of Economic Literature*, 59(1):149–219.

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

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

Hommes, C. H. and Makarewicz, T. (2021). Price level versus inflation targeting under heterogeneous expectations: a laboratory experiment. *Journal of Economic Behavior & Organization*, 182:39–82.

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

Hommes, C. H., Massaro, D., and Weber, M. (2019b). Monetary policy under behavioral expectations: theory and experiment. *European Economic Review*, 118:193–212.

Kahneman, D. and Tversky, A. (1984). Choices, values, and frames. *American psychologist*, 39(4):341.

Kryvtsov, O. and Petersen, L. (2021). Central bank communication that works: Lessons from lab experiments. *Journal of Monetary Economics*, 117:760–780.

Lucas, R. E. (1986). Adaptive behavior and economic theory. *The Journal of Business*, 59(4):S401–S426.

Marimon, R., Spear, S. E., and Sunder, S. (1993). Expectationally driven market volatility: An experimental study. *Journal of Economic Theory*, 61(1):74–103.

Marimon, R. and Sunder, S. (1993). Indeterminacy of equilibria in a hyperinflationary world: Experimental evidence. *Econometrica*, 61:1073–1107.

Marimon, R. and Sunder, S. (1994). Expectations and learning under alternative monetary regimes: An experimental approach. *Economic Theory*, 4:131–162.

Marimon, R. and Sunder, S. (1995). Does a constant money growth rule help stabilize inflation? *Carnegie-Rochester Conference Series on Public Policy*, 43:111–156.

Mauersberger, F. and Nagel, R. (2018). Levels of reasoning in keynesian beauty contests: A generative framework. In Hommes, C. and LeBaron, B., editors, *Heterogeneous agents models*, volume 4 of *Handbook of Computational Economics*, pages 541–634. Elsevier.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, pages 315–335.

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

Pfajfar, D. and Žakelj, B. (2018). Inflation expectations and monetary policy design: Evidence from the laboratory. *Macroeconomic Dynamics*, 22(4):1035–1075.

Sonnemans, J. and Tuinstra, J. (2010). Positive expectations feedback experiments and number guessing games as models of financial markets. *Journal of Economic Psychology*, 31(6):964–984.

Stöckl, T., Huber, J., and Kirchler, M. (2010). Bubble measures in experimental asset markets. *Experimental Economics*, 13(3):284–298.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.