

Bubbles, crashes and information contagion in large-group asset market experiments

Cars Hommes¹, Anita Kopányi-Peuker¹ and Joep Sonnemans¹

¹University of Amsterdam and Tinbergen Institute

August 24, 2018

Abstract

We experimentally investigate how price expectations are formed in a large asset market where subjects' only task is to forecast the future price of a risky asset. The realized prices depend on these expectations. We compare small (6 participants) and large markets (about 100 participants). In large markets the influence of an individual forecast on the realized price is much smaller. When realized prices are far from the equilibrium, a random selection of participants receives “news”.

We observe both stable markets and large bubbles for both small and large markets. The data analysis shows no differences between markets considering group size. Coordination on bubbles is thus robust in large groups. The news has a mitigating effect on individual forecasts and in some markets the news successfully drives prices back towards the fundamental, but we also observe large bubbles in which the news apparently has no effect. The behavioural heuristics switching model with subjects switching to trend-following strategies can capture experimental data well.

JEL classification: C91, C92, D53, D83, D84

Keywords: Experimental finance, expectation formation, learning to forecast, financial bubbles.

Acknowledgements: The authors want to thank for the support of the experimental team at the University of Valencia. We are also grateful for the discussions with the IBSEN-members, and conference participants. Financial support of the H2020 grant of the IBSEN project (“Bridging the gap: from Individual Behavior to the Socio-Technical Man”) is gratefully acknowledged (grant number: 662725).

1 Introduction

Expectations play an important role in everyday life, and in particular in economic activities. Agents form their expectations about the future, and based on these expectations, they decide on what actions to take. For example, traders and producers form price forecasts, and based on these forecasts they buy or sell assets, or decide on how many goods to produce. The actions that agents execute will have an impact on the realised prices. To understand markets we have to understand how people form their expectations. Are they using sophisticated methods? Or simple rules of thumb? Previous research suggest that agents do not necessarily use rational expectations, but rely more on some simple heuristics or learning. For evidence see for example Barberis and Thaler (2003) or Branch and Evans (2010).

One way to investigate expectation formation is through Learning-to-Forecast Experiments (LtFE), first used by Marimon and Sunder (1994). In a LtFE, subjects' only task is to form expectations about the future (e.g. to predict prices, output gap, inflation), and the actions based on the expectations are computerised based on optimal decision making. This way researchers can get clean data on expectations, as subjects are only rewarded based on their forecasting performance. In the past many LtFE were conducted in different economic environments.¹ Hommes (2011) gives an overview of LtF experiments.

Not only LtFE, but macroeconomic experiments in general are more and more common to test policy implications and agents' behaviour in different markets. Duffy (2010) gives an overview of the progress in the field. A common critique to these macroeconomic experiments is about the limited group size. In the real world many more agents are interacting with each other, thus the findings might not be directly applicable outside the lab. With this study we will contribute to this debate by investigating how group size affects expectation formation in a LtFE.

Our design is based on the experiment of Hommes et al. (2008). Subjects need to forecast the price of a risky asset for 50 periods. Their expectations will determine the realised price in the market with a positive feedback between expectations and realisations.² Subjects know this relationship, but they do not know the exact law of motion for the prices. Subjects are paid only based on their forecasting performance. The market price is determined using a mean-variance optimisation of asset trade, which is computerised, and subjects do not need to carry out the trades. In the original paper by Hommes et al. (2008), the authors observed very large bubbles in 5 out of the 6 groups. In their setup 6 subjects formed one market, and the

¹The different frameworks include among others asset pricing experiments (e.g. Hommes et al., 2008), cobweb-models (e.g. Hommes et al., 2007), New-Keynesian macromodels (e.g. Arifovic and Petersen, 2017).

²Note that it is much more difficult to coordinate on the unique rational expectations equilibrium under positive than under negative feedback, especially if the system exhibits a near unit root process. For evidence see Heemeijer et al. (2009), Sonnemans and Tuinstra (2010) or Hanaki et al. (2016).

price depended on the average expectations. Thus, the influence of one subject was relatively high, even though they were not aware of the exact group-size. The natural question arises whether bubbles would still arise in large markets, where each individual has a much smaller weight in determining the market price. In our experiment about 100 subjects form a large market, and we study the price dynamics. In order to have a valid comparison between different group sizes under the same circumstances we also run small markets. In total we have 6 large groups (92-104 subjects in each), and 13 small groups of 6.

In the real world, when markets are (far) out of equilibrium, investors are likely to read in the news discussions about whether the market is in a bubble or not, which may influence their expectations about future prices. This element of real life is typically not incorporated in laboratory experiments. A new feature of our design is that we introduce a news announcement when prices are too high or too low compared to the fundamental. These news elements are short messages that can appear on subjects' screen: "Experts say the stock market is overvalued (undervalued)." The messages do not have a direct effect on the price realisation but might have an effect on subjects' beliefs. Because these messages are received only with a fixed probability we can compare the difference in expectations of participants in the same market who did or did not receive the message.³ In order to prevent prices diverging forever, we also kept an upper bound on the forecasts, but it is high enough to leave room for the news to have an effect. A priori it is not perfectly clear how prices in large markets might look like. One can argue that large markets are much more stable, as each individual has a lower influence on the market, and individual errors are more likely to cancel out (Muth, 1961). On the other hand, bubbles might even be faster in large groups if subjects coordinate on adapting a stronger trend-following behaviour than subjects in smaller groups.

Our experiment shows no substantial difference between small and large groups. We observe both stable and unstable behaviour for both group sizes.⁴ The news on overvaluation successfully influenced subjects' predictions: Those, who see news increase their price expectations in a lesser extent than those not seeing the news (compared to the previous forecast). However, this individual effect is not enough in some markets to prevent very big bubbles to arise which are only stopped by the upper bound on expectations.

Subjects do not have direct information about others' expectations, but nevertheless coordination of expectations is observed, as in Hommes et al. (2008). However, subjects' forecasts are less coordinated in

³In a different context (monetary policy in a macroeconomic experiment) Hommes et al. (2015) provide all participants (bad) news in their "expectational shocks" treatments in beforehand determined specific periods; we let news endogenously arise in our experiment depending only on the price, and only for a subset of participants.

⁴In contrast to Hommes et al. (2008), we have 5 stable small markets out of 13 which is a larger fraction than 1 out of 6. This might simply be due to the existence of possible news.

the large groups when prices rapidly increase along large bubbles than in the small groups.

Even though subjects have all the information needed to calculate the fundamental price, their predictions and the resulting market price are not consistent with a model only containing fundamentalists. The behaviour of about 29% of the subjects is broadly consistent with a first-order heuristics model, in which agents use the last observed price, the last observed price change, their own last prediction, and the sample average to make their decision. Most of these subjects are trend-followers with anchoring on a weighted average of the last observed price, their last prediction and the sample average. We do not find substantial differences between small and large groups for the individual rules either. First-order-heuristics assume that subjects use the same rules in all periods, a more realistic model is that they switch rules when other rules seem to perform better. For this analysis we employed the behavioural heuristics switching model (Anufriev and Hommes, 2012b). This model provides a good fit for our experimental data. We observe that subjects use simple heuristics (rules), and are more likely to switch to another rule if that rule would have performed better in the recent past. We find that in more stable markets the anchoring and adjustment learning rule is dominant, whereas the creation of the bubbles are consistent with a strong trend-following rule.

In experimental economics it is not yet common to run large-scale experiments. Usual labs are simply not large enough for such an experiment, and also it is very costly to pay the standard amount for such a large number of subjects. However, we are not the first to run incentivised large-scale experiments. For example, Isaac et al. (1994), and Weimann et al. (2014) investigated group size effect for public good games. Isaac et al. (1994) considered groups of 4, 10, 40 and 100 in multiple sessions for extra credits rather than cash. They found that large groups provide the public good more efficiently than small groups. Weimann et al. (2014) found similar results by comparing groups of 60 and 100. Gracia-Lázaro et al. (2012) looked at the prisoner’s dilemma game on large networks of about 600 people each. They do not find substantial differences between network types in terms of cooperation rate. Williams and Walker (1993) examined how traders behave in a large-scale double auction with more than 300 traders. In their experiment students were incentivised by extra credits. Bossaerts and Plott (2004) investigated asset markets with about 40 subjects per market. Their main focus was not on group-size however. The most related paper to our study is Bao et al. (2016) who investigated the same setup as Hommes et al. (2008) in 7 groups of 21 to 32 subjects with the usual cash incentives to subjects. In their experiment 6 out of the 7 groups exhibit large bubbles, and these bubbles are created much faster than in the small groups of Hommes et al. (2008). We add to their work by explicitly comparing small and very large groups, and by introducing the news element.

The remainder of this paper is organised as follows. Section 2 describes the experimental economy and

procedures. In section 3 we present the experimental results. Section 4 investigates individual decision rules. In section 5 we describe the heuristics switching model, and its fit on our experimental data. Section 6 concludes.

2 Experimental design

2.1 Experimental market

We consider an asset market with heterogeneous beliefs where agents need to allocate their wealth between a risk-free bond (that pays gross return of $R = 1 + r$) and a risky asset that pays an uncertain dividend with an average dividend of \bar{y} (Campbell et al., 1997; Brock and Hommes, 1998). Based on this allocation decision, the wealth of agent i in period $t + 1$ is given by

$$W_{i,t+1} = RW_{i,t} + z_{i,t}(p_{t+1} + y_{t+1} - Rp_t), \quad (1)$$

where $z_{i,t}$ is the position of the risky asset by agent i in period t (positive or negative) and p_t and p_{t+1} are the prices of the risky asset in periods t and $t + 1$ respectively. We assume that agents maximize a simple myopic mean-variance utility function, that is, they solve the following problem:

$$\max_{z_{i,t}} \left\{ E_{i,t}W_{i,t+1} - \frac{a}{2}V_{i,t}(W_{i,t+1}) \right\} \equiv \max_{z_{i,t}} \left\{ z_{i,t}E_{i,t}\rho_{t+1} - \frac{a}{2}z_{i,t}^2V_{i,t}(\rho_{t+1}) \right\}, \quad (2)$$

where $E_{i,t}$ and $V_{i,t}$ are the individual expectations about wealth and variance. Note that these might not be perfectly rational. Furthermore, a is a parameter for risk aversion, and ρ_{t+1} is the excess return as defined by $\rho_{t+1} \equiv p_{t+1} + y_{t+1} - Rp_t$. For simplicity we assume that the variance is constant and homogeneous across agents, i.e. $V_{i,t}(\rho_{t+1}) = \sigma^2$. Given these assumptions the optimal demand of agent i is given by

$$z_{i,t}^* = \frac{E_{i,t}(\rho_{t+1})}{aV_{i,t}(\rho_{t+1})} = \frac{p_{i,t+1}^e + y_{t+1} - Rp_t}{a\sigma^2}, \quad (3)$$

where $p_{i,t+1}^e = E_{i,t}(p_{t+1})$ is the individual forecast by agent i of the price in period $t + 1$.

The market price for the risky asset is set by market clearing, that is demand equals supply:

$$\sum_{i=1}^N z_{i,t}^* = Z_t^S, \quad (4)$$

where Z_t^S is the exogenous supply. For simplicity, we assume that this exogenous supply is 0. Furthermore we assume a small fraction of noise traders. Their position is incorporated in the equilibrium pricing equation as a small IID noise term, $\varepsilon_t \sim N(0, 0.5)$. This will result in the equilibrium pricing equation

$$p_t = \frac{1}{(1+r)N} \left(\sum_{i=1}^N (E_{i,t}(p_{t+1}) + y) \right) + \varepsilon_t = \frac{1}{(1+r)} (\bar{p}_{t+1}^e + y) + \varepsilon_t,$$

or equivalently

$$p_t = p^f + \frac{1}{R}(\bar{p}_{t+1}^e - p^f) + \varepsilon_t, \quad (5)$$

where \bar{p}_{t+1}^e is the average forecast for period $t + 1$ across all the individuals.

Note that subjects know the realised prices up to $t - 1$ and predict the price of $t + 1$, so two periods ahead, and the average prediction for $t + 1$ determines the price in period t . Also note that the rational expectation equilibrium is that agents expect the price to be equal to the fundamental price, $p^f = \bar{y}/r$. In that case, the law of motion for the price is $p_t = p^f + \varepsilon_t$. To complete the design, we use the following parameters in the experiment: $\bar{y} = 3.3$ and $R = 1.05$. This implies that the fundamental price is 66.

2.2 Implementation

In the experiment subjects are playing the role of a financial advisor of a pension fund. They are informed that they have to forecast the price of a risky asset, and that based on their forecast, the pension funds will have a certain demand of the asset. They are not explicitly told the law of motion (5), but they know that there is a positive feedback in the market (i.e., the higher their forecast is, the higher the expected price will be *ceteris paribus*).

Subjects are only paid on their forecasting performance. Their payoff is determined by the following formula:⁵

$$\text{Payoff}_{i,t+1} = \max \left\{ 0, \left(1300 - \frac{1300}{49} (p_{i,t+1}^e - p_{t+1})^2 \right) \right\}, \quad (6)$$

where $p_{i,t+1}^e$ denotes the forecast of price at period $t + 1$ formulated by subject i in period t without knowing p_t and p_{t+1} is the realised asset price at period $t + 1$. Subjects were informed about this formula, and they were also provided a payoff table showing earnings corresponding to given forecast errors. Subjects accumulated their payoffs during the experiment, which was converted to euros in the end. They received 0.5 euro for each 1300 points they earned.

During the experiment subjects could see past prices, their own actions and whether they received news in a given period, but not others' decisions. As in previous experiments (see e.g. Hommes et al., 2008), we did not explicitly tell subjects the fundamental price, but they would have been able to calculate it as $p^f = \bar{y}/r$.

In order to control huge deviations from the fundamental price, we imposed an upper limit of 1000 on the forecasts. Subjects became only aware of this once they tried to submit a forecast higher than 1000.

⁵The payoff function is the same as in Hommes et al. (2005, 2008) with groups of 6. Bao et al. (2017) compare Learning-to-Forecast versus Learning-to-Optimize treatments where the payoff is based on realized utility. Large bubbles also arise in the Learning-to-Optimize experiments.

They would receive an error message in that case.⁶ Furthermore, we introduced news announcements which stated either “Experts say the stock market is overvalued.” or “Experts say the stock market is undervalued.” This news announcement appeared on screen with some probability when the price was more than 3 times the fundamental, or when it was lower than $\frac{1}{3}$ of the fundamental. When there was news in the market, it appeared only with 25% probability for each subject (independently drawn), and this was common knowledge to the subjects. This way the effect of the news on individual behaviour can be measured, as in each round when the news was depicted, there were some subjects receiving it, and others not. The news appeared until the price was driven back within the given range (with an independent draw for each subject in each round). For an example of the news, see Appendix A. Finally, in order to prevent huge variation in initial predictions, subjects were told that the price in the first period is very likely to be between 0 and 100.

2.3 Treatments and experimental procedure

Large groups around 100 do not fit into a single lab. Therefore, the experiment was conducted by connecting the experimental CREED-lab in Amsterdam and the LINEEX-lab in Valencia via internet. The experiment was programmed in php, and each subject could choose between English and Spanish at the beginning of the session. In the first 6 sessions all subjects formed one large market, whereas in the 7th session subjects were grouped in markets of 6. This results in total in 6 large markets and 13 small markets.⁷ In total 676 (370 in Valencia and 306 in Amsterdam) subjects participated, mainly students with various backgrounds. In order to keep the length of the experiment within a reasonable timespan, a time limit on the decisions was imposed. Subjects had 2 minutes to make a decision in the first 10 periods, and 1 minute in later periods. If subjects did not submit a decision on time, they would not earn anything that period, and the average forecast was determined only based on the other forecasts. Before the experiment, subjects had the opportunity to read the instructions at their own pace from the computer screen, which took on average 40 minutes until everyone completed. The experimental market took on average 1 hour per session. The English instructions are presented in Appendix A (the Spanish instructions are available on request). Subjects earned on average 8.3 euros from the forecasting task (with a minimum of 0 and a maximum of 21.84 euros out of 25 euros) plus a show-up fee, plus additional earnings from an unrelated, surprise one-shot volunteer’s dilemma after the experimental asset market (Kopányi-Peuker, 2018).

⁶Note that this limit is more than 15 times the fundamental price, which allows large bubbles to form.

⁷In the session for the small markets, we had 78 subjects.

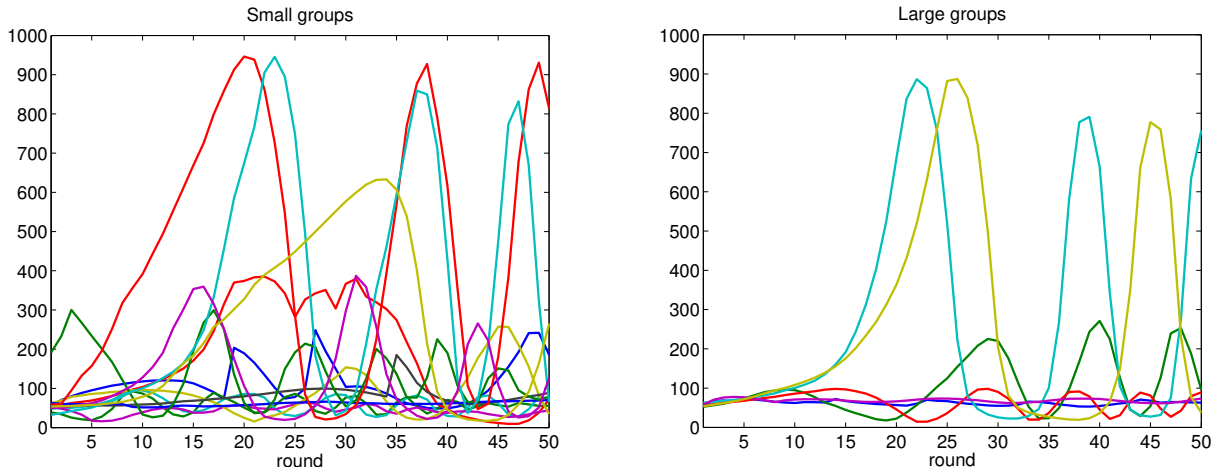


Figure 1: Realized prices in each market for the Small (left panel) and the Large (right panel) treatments

3 Experimental results

In this section we present the results of the experimental asset markets. Section 3.1 describes how the market price evolved in the different markets, and Section 3.2 investigates how individuals coordinated their price forecasts. In Section 3.3 the effect of news on individual behaviour and the market price are discussed. Finally, in Section 3.4 we discuss initial prices and forecasts. Unless otherwise stated all statistical tests are carried out by a two-sided nonparametric test (Mann-Whitney ranksum test or Wilcoxon signrank test).

3.1 Market behaviour

Figure 1 shows the market price for each market in the 13 small groups (left panel) and the 6 large groups (right panel).⁸ The evolution of the market price is very heterogeneous for both the Small and the Large treatments. We distinguish three different qualitative behaviours: (i) some markets are stable or exhibit small oscillations; (ii) other markets have moderately large bubbles (with a peak at about 3-4 times the fundamental price), and finally, (iii) some have very large bubbles, which crashed because subjects reached the highest possible forecast of 1000. Table 1 contains descriptive statistics about each market by presenting the median and mean market price, the standard deviation and the relative (absolute) deviation from the fundamental price. Note that there are some markets which have a relative low average price; these are the more stable. Other markets are heavily overpriced, and they also have rather high standard deviation due to

⁸For one of the large groups, Group 104 with 104 subjects the market price was incorrectly saved. In round 30, 22 of the 104 subjects saw 202.1 instead of 204.1. Note that this difference is negligible based on the fundamental and the group-size. For the data analysis we used the market price of 204.1 for each subject in that round.

the observed bubbles and crashes. The average market price over the 50 periods per treatment is 139.38 for the Large groups and 153.41 for the Small groups (this difference is not statistically significant, $p = 0.93$).⁹ Furthermore, in the Small treatment the average price is significantly higher than the fundamental price of 66 ($p = 0.02$ according to the Wilcoxon signrank test), whereas the average price is not significantly higher than 66 for the Large groups ($p = 0.17$).¹⁰ All markets' standard deviations are significantly higher than the standard deviation predicted by the rational expectation model, that is 0.5, the standard deviation of the noise term.

Table 1 also reports the Relative Absolute Deviation (RAD) and the Relative Deviation (RD) from the fundamental price.¹¹ Most markets are overpriced, and there is no market in which the price never goes under the fundamental price. There is one market in which the price is constantly under the fundamental: group 5 in the Small treatment (group 8 almost does the same). The more stable groups 92, 103 and 8 stay quite close to the fundamental (with relatively low RAD and RD), but the stable group 100-1 oscillates around p^f (with relatively high RAD compared to RD). In markets with bubbles both RAD and RD are relatively high. To sum up, considering only the market price, no striking differences between small and large groups can be observed.¹²

3.2 Coordination of expectations

The next question is, are there differences in individual behaviour in large and small markets? More specifically, whether and how do subjects manage to coordinate their expectations in a market? Figure 2 shows the standard deviation for the individual expectations over time for the small and the large markets. Apparently, there is no trend towards learning to coordinate; we do not observe higher coordination in later periods. The small and large groups differ in the peaks of the standard deviation. High peaks in the small groups correspond mainly to individuals experimenting with different predictions (note that the individual influence is much larger in these groups). However, in the large markets with very large bubbles the

⁹All statistical tests are on the group level ($n = 13$ for the small groups, and $n = 6$ for the large groups), using nonparametric tests.

¹⁰The median prices are significantly higher for both treatments than the fundamental value ($p = 0.046$ for the Large, and $p = 0.03$ for the Small markets).

¹¹These measures are calculated following Stöckl et al. (2010): $RAD = \frac{1}{50} \sum_{t=1}^{50} |p_t - p^f|/p^f$, and $RD = \frac{1}{50} \sum_{t=1}^{50} (p_t - p^f)/p^f$.

¹²Comparing the figures about the marketprice it seems that when there are large bubbles the price increases with a higher growth rate in the large markets. Following Hüsler et al. (2013) we estimate the growth rate of the bubbles, but do not find clear evidence for treatment differences. If anything, the growth rate seems to be faster in the large groups than in the small groups when looking at anchoring on the price. Note however, that we only have 2 small and 2 large markets with very large bubbles, thus our results can only give a suggestive evidence on the issue. This analysis is relegated to Appendix B.1.

group ID	Median	Mean	Std. Dev.	RAD	RD	IE	DE (%)	CE (%)
Group 5	42.77	42.86	11.31	35.06	-35.06	284.51	238.87 (84%)	45.64 (16%)
Group 2	54.89	55.11	19.29	27.79	-16.49	483.81	222.93 (46%)	260.89 (54%)
Group 4	57.04	59.71	20.48	28.14	-9.52	337.00	99.71 (30%)	237.29 (70%)
Group 8	60.47	60.80	4.72	8.72	-7.87	21.99	18.98 (86%)	3.01 (14%)
Group 7	74.84	78.82	24.43	27.21	19.43	1771.99	1458.85 (82%)	313.14 (18%)
Group 6	84.06	90.22	57.85	66.23	36.70	1187.87	568.26 (48%)	619.62 (52%)
Group 12	93.43	131.60	109.91	133.14	99.39	4645.08	1707.59 (37%)	2937.49 (63%)
Group 1	102.80	113.88	51.93	78.31	72.55	6558.04	5264.37 (80%)	1293.67 (20%)
Group 9	126.69	137.01	75.66	118.99	107.59	7535.47	4948.32 (66%)	2587.15 (34%)
Group 3	129.98	177.02	135.13	196.50	168.21	4533.43	3571.60 (79%)	961.83 (21%)
Group 13	173.62	248.54	211.17	297.85	276.57	2806.46	366.89 (13%)	2439.57 (87%)
Group 11	208.93	352.15	308.03	446.66	433.56	27174.41	6084.24 (22%)	21090.17 (78%)
Group 10	392.88	446.54	323.79	588.97	576.58	21681.09	6239.74 (29%)	15441.35 (71%)
Group 92	63.28	62.54	5.18	7.59	-5.24	1318.73	1302.24 (99%)	16.49 (1%)
Group 103	69.11	68.59	4.48	6.54	3.92	438.90	428.74 (98%)	10.16 (2%)
Group 100-1	69.57	64.29	26.15	33.45	-2.60	395.72	111.22 (28%)	284.50 (72%)
Group 99	84.75	104.12	72.22	87.63	57.76	3830.15	2113.17 (55%)	1716.98 (45%)
Group 100-2	106.45	271.20	284.78	333.29	310.90	24772.46	8258.59 (33%)	16513.94 (67%)
Group 104	133.27	265.55	275.71	325.37	302.35	20786.80	8482.93 (41%)	12304.50 (59%)

Notes: Groups 1-13 are the small groups with size 6, and groups 92 to 104 are the large markets. For these markets, the number in the group ID indicates the number of subjects forming that market. Markets are sorted on Median market price. RAD (Relative Absolute Deviation) and RD (Relative Deviation) are in percentages, and are calculated by the average (absolute) deviation from the fundamental divided by the fundamental value. Average individual forecast errors (IE) are calculated by taking the average over individuals and periods of the individual quadratic forecast errors. Dispersion error (DE) is calculated by taking the average of the quadratic differences between individual forecasts and the average forecast for a given period. Finally, common error (CE) is calculated by taking the average of the quadratic difference between the average forecast and the realized price.

Table 1: Descriptive statistics of the markets

heterogeneity between individual forecasts follow the price pattern. Along the bubbles, subjects' forecasts become less coordinated than for lower market prices. The smallest coordination is achieved in the peak of the bubble. After the crash subjects coordinate again better on their forecasts.¹³

To further illustrate coordination between individuals, Figures 3 to 5 plot typical examples of individual predictions over time for both small and large markets; Figure 3 illustrates more stable markets; Figure 4

¹³We find a similar pattern if we plot the ratio of the standard deviation and the mean prediction.

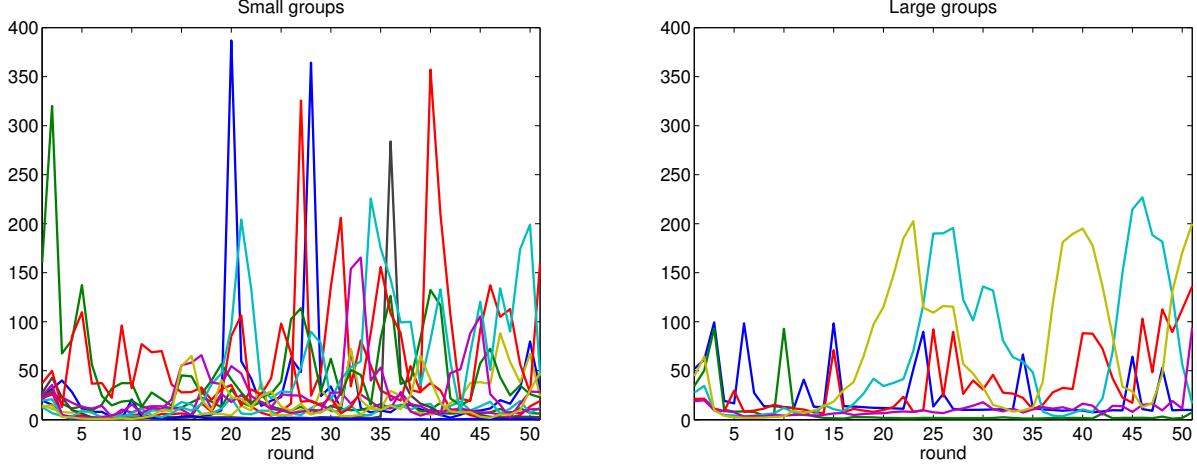


Figure 2: Standard deviation in individual predictions in each market for the Small (left panel) and the Large (right panel) treatments

moderately large bubbles, and Figure 5 very large bubbles.¹⁴ Each black line corresponds to an individual forecast, whereas the realized market price is red. Coordination is apparently harder in the large group, but subjects were quite successful in it for stable and moderately large bubbles. However, in large markets with very large bubbles (Figure 5) coordination becomes lower when the bubble increases and is very weak shortly before the crash. This can also be seen in Figure 2. Subjects tried to decrease the price, but they need a coordinated action to be successful. In the small groups however, subjects coordinated along bubbles as well, not only in stable periods.¹⁵

To quantify the coordination between subjects, we have calculated the average quadratic individual error (IE), the dispersion error (DE) and the common error (CE) for each market. IE is calculated by taking all the individual quadratic forecast errors for each period in which a forecast was submitted, and then we take the average of all these forecasts over individuals and periods for each market ($IE = \frac{1}{N} \sum_{t,i} (p_{i,t}^e - p_t)^2$ where N is the number of individual forecasts in a group over all the 50 periods). This error measures how well subjects predict the market price. The dispersion error is calculated by taking the average of the quadratic difference of the individual forecasts and the average forecast of the given period over all individuals and periods for each market ($DE = \frac{1}{N} \sum_{t,i} (p_{i,t}^e - \bar{p}_t^e)^2$, where N is the number of individual forecasts in a group over all the 50 periods). This variable measures how well subjects coordinate with each other. Finally, the common error (CE) is calculated by taking the average quadratic error of the mean forecast compared to the realized price ($CE = \frac{1}{50} \sum_t (\bar{p}_t^e - p_t)^2$). The common error measures the quality of the average expectations. Table 1 presents IE, DE and CE for each market. By definition DE and CE add

¹⁴The markets are chosen to illustrate each behavioural pattern. Plots for further markets are relegated to Appendix B.2.

¹⁵Cf. with Figure 2 with no peaks at the same time as bubbles for the small groups.

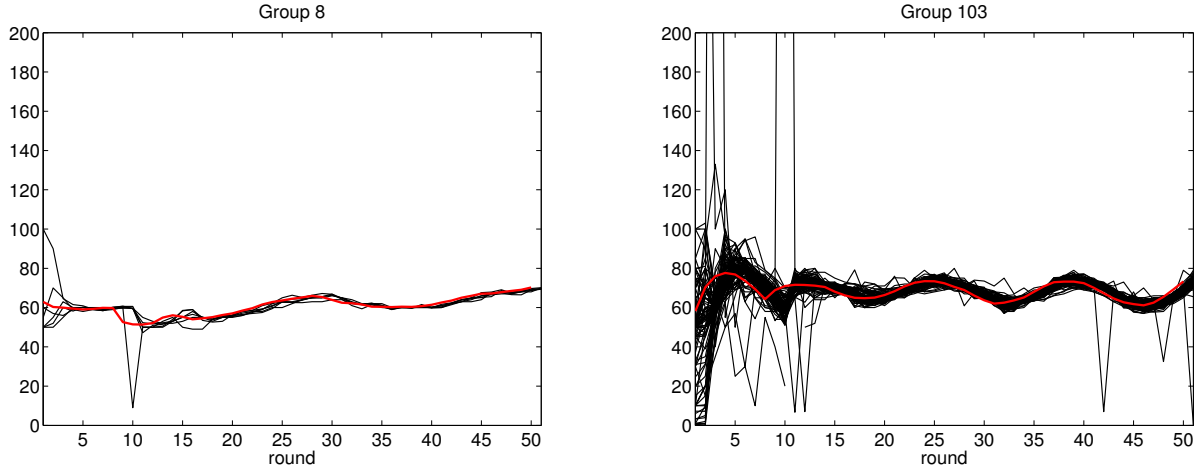


Figure 3: Examples of predictions in stable markets in a small (left) and a large (right) market

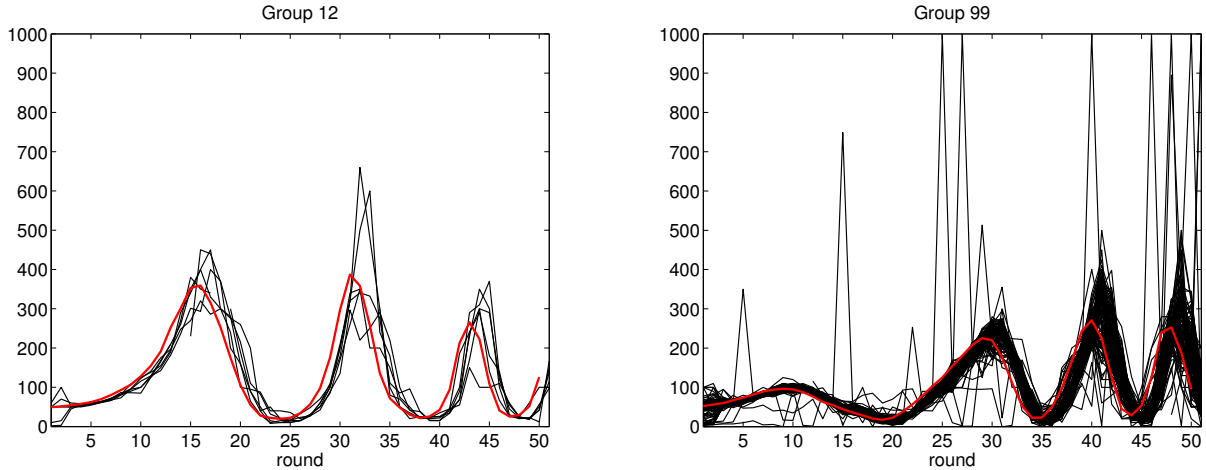


Figure 4: Examples of predictions in moderately large bubbles in a small (left) and a large (right) market

up to IE. Because of this it is easy to see that DE and CE cannot easily be interpreted in absolute terms, but only relative to IE. This is because in a more stable market IE tends to be much smaller than in less stable markets. There is a huge variation in whether the common or the dispersion errors are relatively larger within the individual errors. However, a relatively small common error is more common in the more stable groups. This means that in these markets individual errors tend to cancel out (although of course not perfectly). This happens both in small and large groups. This finding is in line with Muth (1961), who stated that even though individuals make prediction errors, on average they make rational decisions when individual errors are likely to cancel out. In markets with large bubbles CE is much larger than DE. Thus, in these markets aggregate expectations are not rational in the sense of Muth (1961).

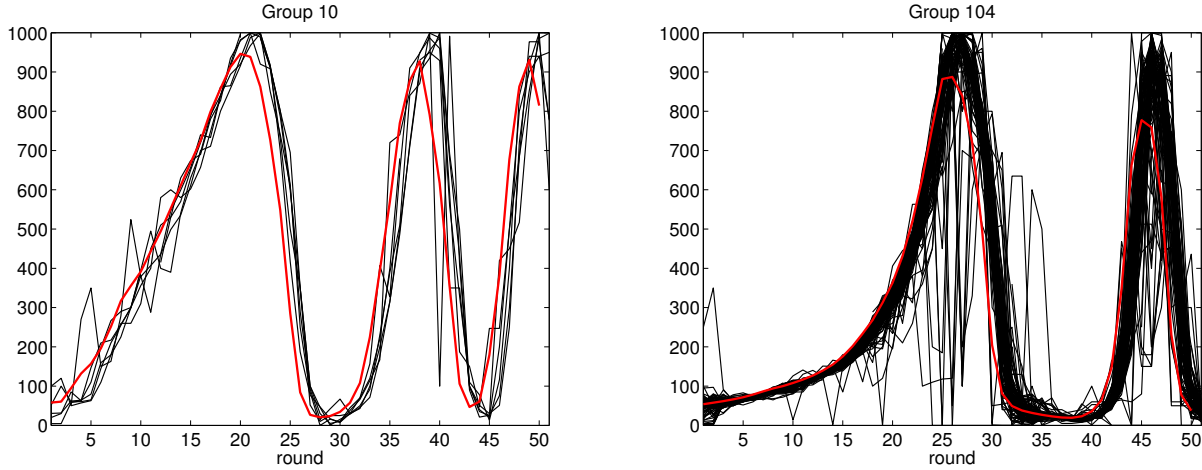


Figure 5: Examples of predictions in very large bubbles in a small (left) and a large (right) market

	1 st bubble	2 nd bubble	3 rd bubble
No news	50.25 (74.05)	235.87 (207.66)	209.59 (226.4)
News	13.4 (60.39)	159.03 (250.56)	100.46 (221.80)
p -value	0.075*	0.16	0.13
# of markets	11	8	7

Notes: *: significant at 10% level according to Wilcoxon-test for the differences within bubbles. Table contains averages over individuals for both small and large markets. Tests are performed on market-level, i.e. we calculate the average change in prediction in each market, thus the number of observation for each test is given by the number of markets in the table.

Table 2: Average increase in prediction after the first news-element is seen in the group for the different bubbles over time - all data

3.3 News announcements

First we discuss the effect of news. News appeared with probability 25% when the last price was sufficiently far away from the fundamental. This analysis regards only the news on overvaluation and disregards the news on undervaluation.¹⁶

News of overvaluation was observed in 8 small groups (from Group 6 onwards in Table 1) and in 3 large groups (from Group 99 onwards in Table 1). Table 2 shows the average increase in prediction after the first news element is seen in a group, for each bubble separately for those who have seen the news, and those who have not seen the news.¹⁷ The data is pooled for small and large markets as there are no significant

¹⁶News of undervaluation was observed 9+20 times only in total (in large and small groups, resp.), whereas news of overvaluation was observed 48+136 times (in large and small groups, resp.).

¹⁷Only the first period in which news was displayed for at least one member of the group, is included in the analysis, because differences in later periods might be influenced by potential views of the news in earlier periods. There was only one

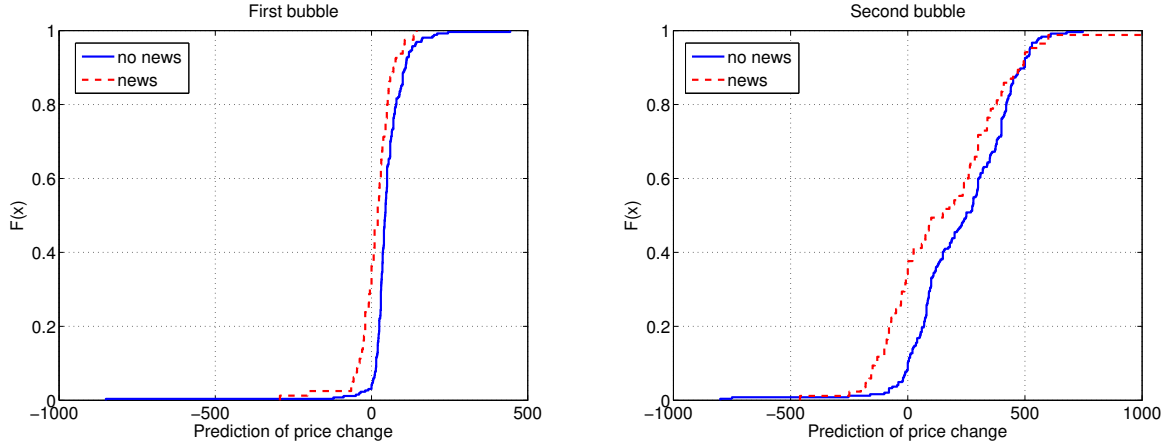


Figure 6: Cumulative distribution function of predicted price changes depending on whether the news has been seen, along the first (left figure) and second (right figure) bubbles

differences across treatments in any of these cells ($p > 0.17$). Subjects who saw the news in the first bubble increased their predictions with a significantly lower amount than those who have not seen the news ($p = 0.075$). A similar decrease in predictions after news is seen for the second and third bubbles, but the differences are not statistically significant. On average, subjects who saw the news still followed the upward price-trend (the average predicted increase is positive), but in a more moderate way. Figure 6 shows the cumulative distribution function of the predicted price change for those who have seen the news and for those who have not. Here we only focus on the first two bubbles, as the third bubble looks about the same as the second. In both figures subjects report lower predicted price changes if they have seen the news. Another striking feature of Table 2 is that the average increase in predictions substantially grows from bubble nr. 1 to bubble nr. 2. It looks as if the first bubble is generally slower than later bubbles when subjects already gained experience with bubbles, crashes, and the news element. However, these differences are not statistically significant at conventional levels.

3.4 Initial prices and forecasts

Another question is whether behaviour is path-dependent: is there any difference between initial conditions for markets which exhibit more stable behaviour and markets which have large bubbles? To answer this question, we have classified the markets in two categories depending on whether they are more stable, or exhibit bubbles.¹⁸ Table 3 shows the average initial predictions for periods 1-3, and the initial prices

market with 5 repeated bubbles, these last two bubbles are excluded from the analysis because of insufficient observations.

¹⁸We classified markets without any news of overpricing as more stable markets, and markets, where news was triggered by deviations from the fundamental as markets with bubbles. We decided to use only two different categories as we are

	\bar{p}_1^e	\bar{p}_2^e	\bar{p}_3^e	p_1	p_2
Stable markets	49.50 (35.69)	55.20 (45.03)	64.47 (74.97)	53.54 (7.21)	54.98 (11.70)
Bubble markets	49.92 (38.28)	57.40 (60.15)	62.83 (27.78)	65.61 (42.56)	73.83 (53.52)
p -value	0.75	0.86	0.36	0.80	0.36

Notes: Table contains averages over individuals for predictions for pooling large and small markets. Tests on predictions in period 1 and 2 are performed on individual level (as no interaction between group-members were made at that point), other tests are performed on market-level with 8 stable (3 large and 5 small) and 11 unstable (3 large and) markets.

Table 3: Average initial predictions and prices in the first periods

for periods 1-2. The markets with stable and unstable behaviour show about the same expectations and prices in the first three periods.¹⁹ The difference in market price starts to be significant only in period 4 (thus, based on the prediction for period 5, not in the table). This seems to suggest that the difference between stable and unstable markets is not in the initial conditions.

Note that subjects had to submit expectations for the first two periods at the same time, that is, they do not have additional information for submitting expectation for period 2 compared to period 1. It is striking that subjects apparently expected a positive trend in prices from the beginning. The first period expectations do not significantly differ from 50 for either category which is the midpoint of the (0,100) interval suggested as the most likely prices in the first period. However, if we look at the second period expectations, we see that these are significantly larger than 50 for both stable and unstable groups (both p -values are 0.00).

4 Individual forecasting strategies

For 50 periods subjects made forecasting decisions, and in any period, they only knew the past realised prices and their own previous forecasts. In this section we estimate the individual decision rules. Even though there are only two types of information (past prices and own forecasts) available for subjects, there are many different behavioural rules that could play a role in decision making. Subjects could use different number of lags of the above-mentioned variables, and weight them differently. To restrict our analysis, we will focus here on the first-order anchor and adjustment heuristics (see e.g. Heemeijer et al., 2009 or Bao et al., 2016) which is given by the following equation:

$$p_{i,t+1}^e = \alpha_1 p_{t-1} + \alpha_2 p_{i,t}^e + (1 - \alpha_1 - \alpha_2) \bar{p} + \beta(p_{t-1} - p_{t-2}) + v_t, \quad (7)$$

interested in whether bubbles start because of different initialisations.

¹⁹The marketprice seems to be higher in the bubble markets because of one small market: one of the individuals chose about 500-600 for the first two predictions.

where $\bar{p} = \frac{1}{50} \sum_{t=1}^{50} p_t$ is the average market price over the periods. Instead of \bar{p} Heemeijer et al. (2009) included the fundamental price, as subjects in that experiment had all the information to compute this. However, as we can see in Table 1, the price is on average in many markets much higher than the fundamental. Because of this, it is more realistic to assume that subjects try to learn this price rather than the fundamental, thus, similarly as Bao et al. (2016) we decided to include the average price over the 50 periods as a possible anchor for the subjects.²⁰ This first-order heuristics forecasting rule is relatively simple but can capture different heuristics. For $\alpha_1 + \alpha_2 = 1$ and $\beta = 0$ the rule reduces to adaptive expectations. Taking $\alpha_1 = 1$, and $\alpha_2 = \beta = 0$ we get the naive expectations rule. Finally, if all the parameters are 0 ($\alpha_i = \beta = 0$), then expectations are described by the sample average price. If we consider this price as the ‘equilibrium’-price, then this case can be characterised by a variant of the ‘fundamental expectations’. More generally, the rule is in the form of an anchoring and adjusting rule (see also Section 5) with an anchor using the last observed market price, own forecast, and sample average. This anchor is adjusted in every period based on the last price change; $\beta > 0$ corresponds to trend-following behaviour, whereas $\beta < 0$ represents contrarian behaviour).

For each individual equation (7) is estimated, after removing outliers and filling up missing data (using linear interpolation).²¹ The first 5 periods are disregarded, to allow for a short learning phase. We estimate the model twice for each individual. Once we estimate the given model without removing insignificant regressors. Second, we remove stepwise the variables which are not significant at 5%-level. Besides estimating the parameters, in this latter case we also looked at whether the final model has autocorrelation, heteroskedasticity in the errors, or is possibly misspecified (Ramsey RESET test). The models of 216 of the 676 subjects survive all three model-specification tests at the 5%-level.

Table 4 summarises the average coefficients of the regressions. Panel A presents the average coefficients over individuals from the first, simple estimation. Thus, here we average over all the individuals no matter whether the corresponding coefficient is significantly different from 0. Panel B presents the average coeffi-

²⁰Note that this is a simplified assumption, given that this sample average is not available for subjects at the time they make their decision. We could have also include the sample average of the last *observed* prices, as in the LAA heuristic in Section 5 which usually quickly coverges to the average market price over all the periods. Also, we do not considers returns as in Section 3.1 in the regression, because while subjects can easily get a sense of the average price by looking at the previous prices, they cannot easily estimate returns. Thus, even though we could include it in the regression, it is very unlikely that subjects use that information for their decision making.

²¹An observation was considered as an outlier if the forecast change was higher than 100% of the previously observed price (or in case of low prices, more than 200%). Furthermore, outliers are judged individually (e.g. in case of a structural break, no outlier). In total, of the $676 \cdot 50 = 30,800$ decisions, we have only 56 outliers by 54 subjects (0.18%). Furthermore, from period 3 onwards we had 204 missing forecasts we added by linear interpolation (0.66% of all decisions).

Market type	fraction of subject	$\bar{\alpha}_1$	$\bar{\alpha}_2$	β
<i>Panel A: All models</i>				
Small pooled		0.49 (0.37)	0.31 (0.30)	0.80 (0.36)
Small stable		0.40 (0.38)	0.31 (0.28)	0.65 (0.38)
Small bubble		0.54 (0.35)	0.31 (0.32)	0.90 (0.31)
Large pooled		-0.16 (1.22)	0.58 (0.31)	0.91 (0.36)
Large stable		0.29 (0.29)	0.56 (0.19)	0.88 (0.44)
Large bubble		-0.61 (1.57)	0.59 (0.40)	0.94 (0.26)
<i>Panel B: Good models - significant coefficients</i>				
Small pooled	40% (31/78)	0.70 (21)	0.52 (19)	0.84 (30)
Small stable	40% (12/30)	0.76 (8)	0.43 (6)	0.79 (11)
Small bubble	40% (19/48)	0.67 (13)	0.57 (13)	0.87 (19)
Large pooled	31% (185/598)	0.64 (54)	0.66 (161)	0.97 (172)
Large stable	45% (132/295)	0.60 (41)	0.62 (118)	1.00 (121)
Large bubble	17% (53/303)	0.79 (13)	0.76 (43)	0.89 (51)

Notes: Markets are divided into stable and bubble markets in the same way as in Table 3. In Panel A averages are taken over all individuals, whereas in Panel B only over individuals who had a good model, and whose parameter is significantly different from 0. In Panel A the numbers in brackets indicate the standard deviation, whereas in Panel B the number of observation taken for the averages.

Table 4: Summary of first-order heuristics

cients of the good models which are significantly different from 0. Here also the fraction of subjects having a ‘good model’ is displayed. Typically, subjects seem to take into account both their own last forecast, and the last observed price. Furthermore, on average they seem to be trend following, with $0 < \beta < 1$. The estimated coefficients are similar in the different markets. We do not find significant differences except for the α_2 parameter: considering all models, this parameter is significantly larger in the large groups ($p = 0.03$), and weakly significantly large in the large stable groups than in the small stable groups ($p = 0.053$). Considering only the good models, α_2 is significantly larger in the large stable markets than in the small stable markets ($p = 0.03$). This means that in these markets subjects give a higher weight to their own last prediction than in the small markets. Furthermore, overall, subjects seem to be more trend-following in the small unstable markets, than in the small stable markets ($p = 0.04$), however this difference vanishes when considered only the good models ($p = 0.57$). All the other differences are insignificant at the 10%-level. In the small groups 31 out of the 78 subjects’ behaviour can be described by first-order heuristics. These

31 subjects are distributed among 11 groups. These groups consist of both stable and unstable markets. 8 out of the 31 participants have $\alpha_1 > 0.75$ and $\alpha_2 = 0$ which suggest an anchor that is mainly based on the last observed price (corresponding to a naive expectation if $\beta = 0$). All but one subject has a significantly positive β parameter, ranging from 0.21 (Gr. 7) to 1.49 (Gr. 5). Both these extrema are observed in stable markets. 6 subjects have a pure trend-following rule with $\alpha_1 \in (0.9, 1.1)$, $\alpha_2 = 0$ and $\beta > 0$. A similar pattern is found in the 6 large groups. There is no clear difference between stable and unstable markets. A smaller fraction of the people seems to use a naive anchor compared to the small group, as only 15 out of the 185 subjects have $\alpha_1 > 0.75$ and $\alpha_2 = 0$. 5 subjects seem to have an adaptive anchor with $|1 - \alpha_1 - \alpha_2| < 0.01$. 13 subjects have an insignificant β coefficient, all but one other individuals are trend-following with β ranging from 0.25 (Gr. 92) to 2.05 (Gr.103). One subjects is estimated to be a contrarian with $\beta = -0.17$ (Gr. 92). All of these extrema are found in the more stable markets. Interestingly we find three large groups (one more stable, and 2 less stable markets) in which subjects do not seem to use the last observed price as an anchor. Note however, that they still react on the last price change, and also to the average market price, so they do use the price information.

To summarise this section, about 32% of our subjects' behaviour can be described by the first-order heuristics given by Eq. (7). The estimation results suggest that subjects are mainly trend-followers, but they also use some anchor to base their decision on. This anchor is most of the time mixed, consisting of the last observed price, the last own forecast, and the sample average. No substantial differences between small and large groups are observed. Note that by estimating the first-order heuristics, we restrict our subjects to only use one rule for the whole experiment. However, subjects might change the rules they use over time.²² These switches cannot be described by this simple rule and are the topic of the next section.

5 Heuristic switching model

5.1 Model setup

As the previous sections have shown, subjects do not use homogeneous rational expectations when forming their forecast. It has been argued before that subjects in learning-to-forecast experiments are bounded rational and base their decisions on different rules of thumb (e.g. Hommes et al., 2008). However, subjects are also willing to evaluate these rules again and again, and change them if they yield better forecasts. This behaviour is captured and modelled in the Heuristics Switching Model (HSM) by Anufriev and Hommes (2012b) which is an extension of Brock and Hommes (1997). The idea behind the model is that agents do

²²In the large groups there were some subjects who clearly changed their strategy over time, e.g. by being trend-following for some time, and then reverting to very low predictions in case of large bubbles (e.g. subject 39 in Group 104).

not use a single forecasting rule, but they are heterogeneous in the rules they are using. Furthermore, they also switch between these rules, based on their relative performance. The HSM has already described well experimental data of other learning-to-forecast experiments (e.g. Anufriev and Hommes, 2012a).

To be more precise, we follow Anufriev and Hommes (2012b) for modelling the behaviour by a HSM. Agents are assumed to use four different heuristics, and the average expected price in the market equals the weighted average of the expected prices produced by the heuristics: $\bar{p}_{t+1}^e = \sum_{i=1}^4 n_{i,t} p_{i,t+1}^e$ where $n_{i,t}$ is the fraction of agents using heuristics i in period t . This average expectation is used then in (5) to calculate the realised price. It is important to note that the rules used in HSM only uses information that is available for subjects in the experiment. That is, expectations are formed based on last realised prices, and previous forecasts. In particular, the following four rules were used:²³

$$\begin{aligned}
\textbf{Adaptive expectations:} \quad & p_{ADA,t+1}^e = 0.65 p_{t-1} + 0.35 p_{1,t}^e, \\
\textbf{Weak trend-following rule:} \quad & p_{WTR,t+1}^e = p_{t-1} + 0.4 (p_{t-1} - p_{t-2}), \\
\textbf{Strong trend-following rule:} \quad & p_{STR,t+1}^e = p_{t-1} + 1.3 (p_{t-1} - p_{t-2}), \\
\textbf{Anchor \& adjustment rule:} \quad & p_{LAA,t+1}^e = 0.5 (p_{t-1}^{av} + p_{t-1}) + (p_{t-1} - p_{t-2}),
\end{aligned}$$

where p_{t-1}^{av} is the sample average of realised prices in the last t_1 periods.

These four rules lead to different types of aggregate behavior. Under adaptive expectations prices converge (slowly) monotonically to the fundamental price. Under weak trend-following rule small price trends occur with some minor over- and undershooting, but in the medium to long run price converges to the fundamental. Under the strong trend-following rule the market is unstable and a large bubble occurs. Finally, under the anchor and adjustment rule prices exhibit persistent oscillatory behavior. This is due to the flexible anchor of this rule which gives 50% weight to the average price p_{t-1}^{av} (a proxy for the long run equilibrium price). The anchor and adjustment rule is the only rule able to predict turning points, thus supporting oscillatory behavior.

As agents are able to switch between these forecasting rules, we also need to specify the process agents use to switch between the rules. In every period the decision rules are evaluated by a performance measure (U_i) that corresponds to how subjects are paid in the experiment: $U_{i,t-1} = -(p_{t-1} - p_{i,t-1}^e)^2 + \eta U_{i,t-2}$, where $\eta \in [0, 1]$ is the strength of agents' memory. If η is high, then agents remember the past performance better, whereas for small η agents give a higher weight to the current performance. Based on this performance measure, the evolution of fraction follows the discrete choice model with some inertia. In each period, a fraction δ of agents does not switch rules, whereas $1 - \delta$ fractions follow the evolutionary

²³Parameters are calibrated from earlier experimental results (Anufriev and Hommes, 2012b).

selection based on past performance. This gives the following law of motion for the fractions over time:

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

where β is the intensity of choice parameter. When $\beta = 0$, then all fractions are equal, and the performance does not matter. The higher β is, the more likely is that a better rule gets selected. If $\beta = +\infty$, agents who update their strategy always switch to the best performing rule. Note that this model assumes that agents are able and willing to calculate the performance of all rules, even if a rule is not used in a given period.

5.2 Simulations of experimental markets

In order to look at the model performance on the experimental data, one-period ahead simulations are run using the observed experimental prices. We simulate the model by using previous forecasts generated by the model and past experimental prices. We simulate the prices by the model as well, but these are only used to compare them with experimental prices and are not fed back into the model and into the forecasting rules. It is important to note that the model only uses the same information as subjects had in the experiment.²⁴ To initialize the model, the first two prices in the experiment are used. Furthermore, as Adaptive Expectations rule uses past forecasts as well, we initialize that rule by fixing $p_{ADA,3}^e = 50$ which was the midpoint of the interval we gave the subjects as a very likely price realisation for the first two periods. Furthermore, we fixed the initial shares $n_{i,3} = 0.25$ and $U_{i,3} = 0$ for each decision rule. Following Anufriev and Hommes (2012b), parameters were fixed at $\beta = 0.4$, $\eta = 0.7$, and $\delta = 0.9$.²⁵

Figure 7 shows the average fractions of the different rules for both the small and the large markets. These plots are quite similar across group size. The share of the two trend-following rules follow about the same pattern for both small and large groups, whereas after round 30, there is a difference in the shares of the Adaptive Expectations and Anchor and Adjustment rules between group-sizes.

Figures 8 and 9 show the average fractions in stable and unstable markets (again with unstable markets defined as those receiving news of overvaluation). The simulated fraction for the unstable markets follow about the same pattern for both small and large groups. In unstable markets the Strong Trend-following rule dominates the market first, and when crashes happen, the Anchor and Adjustment rule gains fraction,

²⁴Note that we do not incorporate the possible news in the rules. However, if news has an effect, it might be captured by a higher switching intensity between rules.

²⁵The results reported here are fairly robust w.r.t. the parameters β , η and δ . Furthermore, the results are robust against changes in the coefficients of the 4 rules as long as these changes do not affect the qualitative behavior of each of the rules as described above.

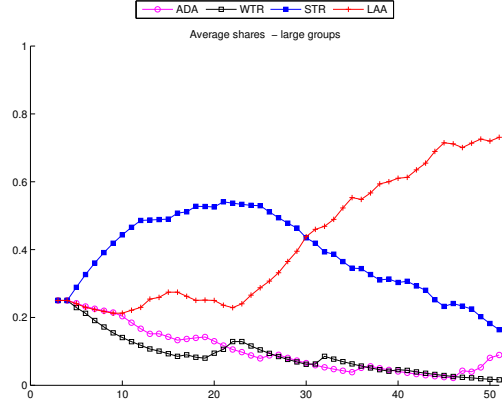
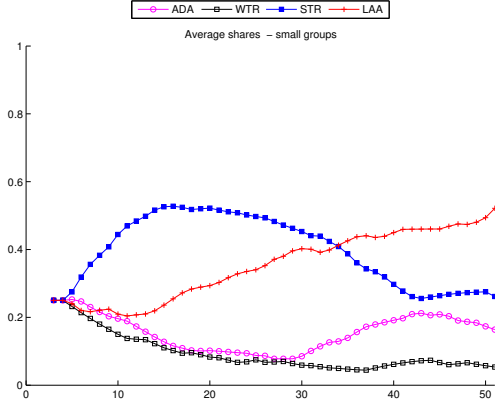


Figure 7: Average simulated fraction of rules for small (left) and large (right) markets

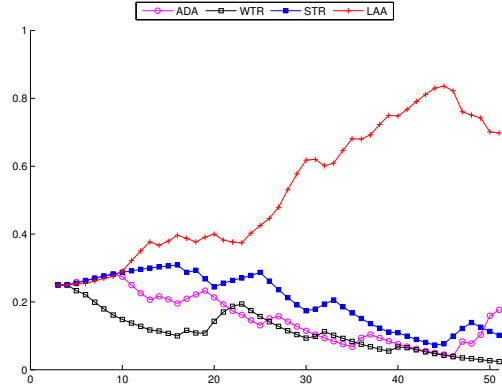
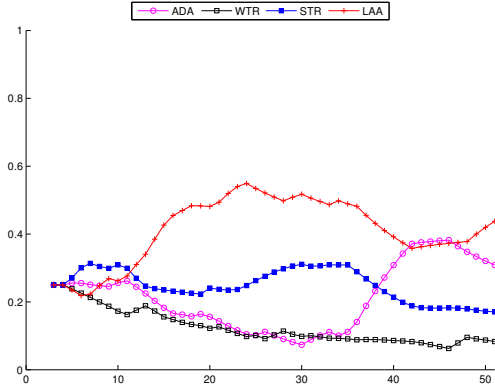


Figure 8: Average simulated fraction of rules for stable small (left) and large (right) markets

and the Strong Trend-following rule's weight decreases after round 20. The dominance of the Strong Trend-following rule leads to bubbles in these markets arising due to coordination on the strong trend-following rule. Expectations are optimistic, and also self-fulfilling. However, prices increase in a much faster rate than subjects predict it, thus along the bubbles they earn very little. Expectations are then reversed either by the depicted news, or by reaching the upper bound, and subjects' behaviour is more in line with the Anchor and Adjustment rule. In the more stable markets there are some differences between small and large groups (Figure 8). Note however, that in the more stable markets all rules perform similarly, thus it is hard to draw strong conclusions here. What is common for both group-sizes is that the Anchor and Adjustment rule is dominant. This rule can more easily capture price oscillations and is also able to describe the crashes. Thus, overall, we do not find substantial differences between small and large groups with the Heuristic Switching Model either.

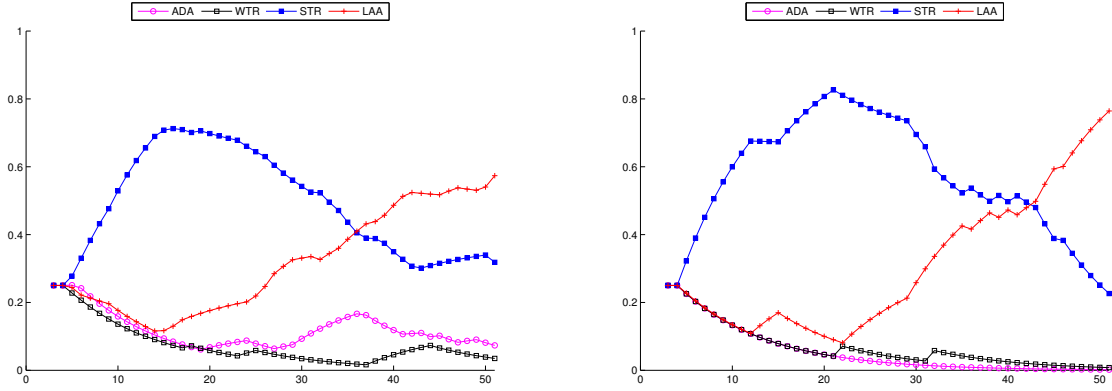


Figure 9: Average simulated fraction of rules for unstable small (left) and large (right) markets

6 Conclusion

A common objection to macroeconomic experiments is that laboratory markets are small and thus the decisions of one individual can have a stronger influence on the outcomes than in real markets. In this study we compare expectation formation in large (about 100 participants) and small (6 participants) experimental markets. Furthermore, we add realism to the markets by providing news from experts about the market. When the price is far away from the fundamental, we randomly choose some of the market participants to receive news about the market being either over- or undervalued. The question is whether these news incidents will be successful in driving prices back to the fundamental, and crash bubbles?

Our experiment reveals no substantial difference between small and large groups. In both cases we can observe relatively stable markets, markets with large bubbles and markets with very large bubbles. The news did not stabilise prices in all markets, as some bubbles crashed because of the artificial upper bound we imposed on prices. However, subjects reacted on news; those who have seen the news of overvaluation have a significantly lower price change compared to those who have not seen the news.

To shed more light on individual behaviour, we estimated individual decision rules by means of the first-order heuristics. This rule is based on the last observed price, last own forecast, last price change and the sample average. Our results suggest that subjects' behaviour is broadly consistent with an anchoring and adjustment rule, where the anchor consists of the last observed price, the sample average and the last own forecast. This anchor is the adjusted in the direction of the last price change, that is according to a trend-following rule. We do not find substantial differences between group-sizes.

To further investigate individual behaviour, we fitted the heuristic switching model on the observed data. This model uses four simple rules that subjects might use in the experiment to form their expectations.

These rules include Adaptive Learning, Weak and Strong Trend-following rules, and Anchoring and Adjustment learning rule. Simulating the heuristic switching model one-period ahead results in a price-series that follows the experimental data quite well. For markets that exhibit large bubbles, the Strong Trend-followers dominate. This rule performs the best in an environment where prices increase rapidly, which can lead to a majority using this rule. This again increases prices, the process repeats itself, and bubbles are unavoidable. However, once the bubbles burst, the Anchoring and Adjustment rule starts to gain ‘users’, and the Strong Trend-following rule loses fractions. In the more stable markets the four rules perform very similarly, as the price is much closer to the fundamental value. In those cases, there is no dominating rule, fractions are more evenly distributed.

To sum up, we have investigated whether well-established results in the learning-to-forecast literature from small groups are also valid in large groups. We found that in this context group-size does not matter. However, this result cannot be extrapolated to all macroeconomic experiments; more research is needed in other contexts.

References

- Anufriev, M. and Hommes, C. (2012a). Evolution of market heuristics. *The Knowledge Engineering Review*, 27(02):255–271.
- Anufriev, M. and Hommes, C. (2012b). 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.
- Bao, T., Hennequin, M., Hommes, C., and Massaro, D. (2016). Coordination on bubbles in large-group asset pricing experiments. *CeNDEF Working Paper, University of Amsterdam*.
- Bao, T., Hommes, C., and Makarewicz, T. (2017). Bubble formation and (in) efficient markets in learning-to-forecast and optimise experiments. *The Economic Journal*, 127(605):F581–F609.
- Barberis, N. and Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, 1:1053–1128.
- Bossaerts, P. and Plott, C. (2004). Basic Principles of Asset Pricing Theory: Evidence from Large-Scale Experimental Financial Markets. *Review of Finance*, 8(2):135–169.

- Branch, W. A. and Evans, G. W. (2010). Asset return dynamics and learning. *The Review of Financial Studies*, 23(4):1651–1680.
- Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica: Journal of the Econometric Society*, pages 1059–1095.
- Brock, W. A. and Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic dynamics and Control*, 22(8):1235–1274.
- Campbell, J. Y., Lo, A. W., MacKinlay, A. C., et al. (1997). *The econometrics of financial markets*, volume 2. princeton University press Princeton, NJ.
- Duffy, J. (2010). Experimental macroeconomics. In *Behavioural and Experimental Economics*, pages 113–119. Springer.
- Gracia-Lázaro, C., Ferrer, A., Ruiz, G., Tarancón, A., Cuesta, J. A., Sánchez, A., and Moreno, Y. (2012). Heterogeneous networks do not promote cooperation when humans play a prisoner’s dilemma. *Proceedings of the National Academy of Sciences*, 109(32):12922–12926.
- Hanaki, N., Sutan, A., and Willinger, M. (2016). The strategic environment effect in beauty contest games.
- 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. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic dynamics and control*, 35(1):1–24.
- Hommes, C., Massaro, D., and Salle, I. (2015). Monetary and fiscal policy design at the zero lower bound—evidence from the lab. *CeNDEF Working Paper 15-11, University of Amsterdam*.
- Hommes, C., Sonnemans, J., Tuinstra, J., and Van de Velden, H. (2005). Coordination of expectations in asset pricing experiments. *Review of Financial studies*, 18(3):955–980.
- Hommes, C., Sonnemans, J., Tuinstra, J., and Van De Velden, H. (2007). Learning in cobweb experiments. *Macroeconomic Dynamics*, 11(S1):8–33.
- Hommes, C., Sonnemans, J., Tuinstra, J., and van de Velden, H. (2008). Expectations and bubbles in asset pricing experiments. *Journal of Economic Behavior & Organization*, 67:116–133.

- Hüsler, A., Sornette, D., and Hommes, C. H. (2013). Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price. *Journal of Economic Behavior & Organization*, 92:304–316.
- Isaac, R. M., Walker, J. M., and Williams, A. W. (1994). Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups. *Journal of Public Economics*, 54(1):1–36.
- Kopányi-Peuker, A. (2018). Yes, I’ll do it: A large-scale experiment on the volunteer’s dilemma. *CeNDEF Working Paper 18-01*.
- Marimon, R. and Sunder, S. (1994). Expectations and learning under alternative monetary regimes: an experimental approach. *Economic Theory*, 4(1):131–162.
- Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica*, 29(3):315–335.
- 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.
- Weimann, J., Brosig-Koch, J., Heinrich, T., Hennig-Schmidt, H., Keser, C., and Stahr, C. (2014). An explanation of (first round) contributions in public-good experiments. *CESifo Working Paper Series*, (5039).
- Williams, A. W. and Walker, J. M. (1993). Computerized laboratory exercises for microeconomics education: Three applications motivated by experimental economics. *The Journal of Economic Education*, 24(4):291–315.

Appendix A Experimental instructions

In this appendix an example of the news (see Figure 10) and the experimental instructions are presented. The difference in the two treatments is only in the information about the groupsize (see *italic*). After the instructions we present the payoff table subjects received on their desks.



Figure 10: News-item when the stock is overvalued

PAGE 1

Welcome to this experiment on decision-making. Please read the following instructions carefully. If you have any questions, please raise your hand, and we will come to your table to answer your question in private.

General information

You are a **financial advisor** to a pension fund that wants to optimally invest a large amount of money. The pension fund has two investment options: a risk free investment (on a bank account) and a risky investment (on the stock market). As their financial advisor, you have to predict the stock price during 51 subsequent time periods. The more accurate your predictions are, the higher your total earnings are.

Forecasting task of the financial advisor

Your only task is to forecast the stock price in each time period as accurate as possible. The stock price has to be predicted **two** time periods ahead. At the beginning of the experiment, you have to predict the stock price in the first two periods. It is very likely that the stock price will be between 0 and 100 in the first two periods. After all participants have given their predictions for the first two periods, the stock price for the first period will be revealed and, based upon your forecasting error, your earnings for period 1 will be given. After that you have to give your prediction for the stock price in the third period. After all participants have given their predictions for period 3, the stock price in the second period will be revealed and, based upon your forecasting error, your earnings for period 2 will be given. This process continues for in total 51 time periods.

The available information for forecasting the stock price in period t consists of

- all past prices up to period $t - 2$, and
- your past predictions up to period $t - 1$, and
- total earnings up to period $t - 2$

In each round you have enough, but limited time to make your forecasting decision. If you do not submit a forecast during this time frame, your pension fund will be inactive, and you will not earn any points in that given round. A timer will show you the remaining time for each period (2 min in the first 10 periods, 1 min in the later periods).

Information about the stock market

The stock price in period t will be that price for which aggregate demand equals supply. The supply of stocks is fixed during the experiment. The demand for stocks is determined by the aggregate demand of a number of large pension funds active in the market. The higher the average demand for stocks is, the higher the realized price will be on the market. There are about $100 / 6$ pension funds in the stock market. Each pension fund is advised by a participant of the experiment.

PAGE 2

News

Throughout the experiment you might receive news from financial experts about the state of the stock market. Examples of news are:

“Experts say the stock market is overvalued.”

“Experts say the stock market is undervalued.”

The news has no direct effect on the stock market, but may affect price predictions of financial advisors. When there is news, on average only 1 out of 4 subjects will receive news. Note that it is also possible that you do not receive any news during the 51 periods.

Earnings

Your earnings depend only on the accuracy of your predictions. The maximum possible points you can earn for each period (if you make no prediction error) is 1300, and the larger your prediction error is, the fewer points you can make. You will earn 0 points if your prediction error is larger than 7. There is a Payoff Table on your desk, which shows the points you can earn for different prediction errors.

We will pay you in cash at the end of the experiment based on the points you earned. You earn 0.5 euro for each 1300 points you make plus an additional 5 euros of participation fee.

Information about the investment strategies of the pension funds

The precise investment strategy of the pension fund that you are advising and the investment strategies of the other pension funds are unknown. The bank account of the risk free investment pays a fixed interest rate of 5% per time period. The stock pays an uncertain dividend in each time period. Economic experts have computed that the average dividend is 3.3 euro per period. The realized stock return per period is uncertain and depends upon the (unknown) dividend and upon stock price changes. Based upon your stock price forecast, your pension fund will make an optimal investment decision. The higher your price forecast is, the more money will be invested in the stock market by the fund, so the larger will be their demand for stocks.

On the next screens you are asked to answer some understanding questions.

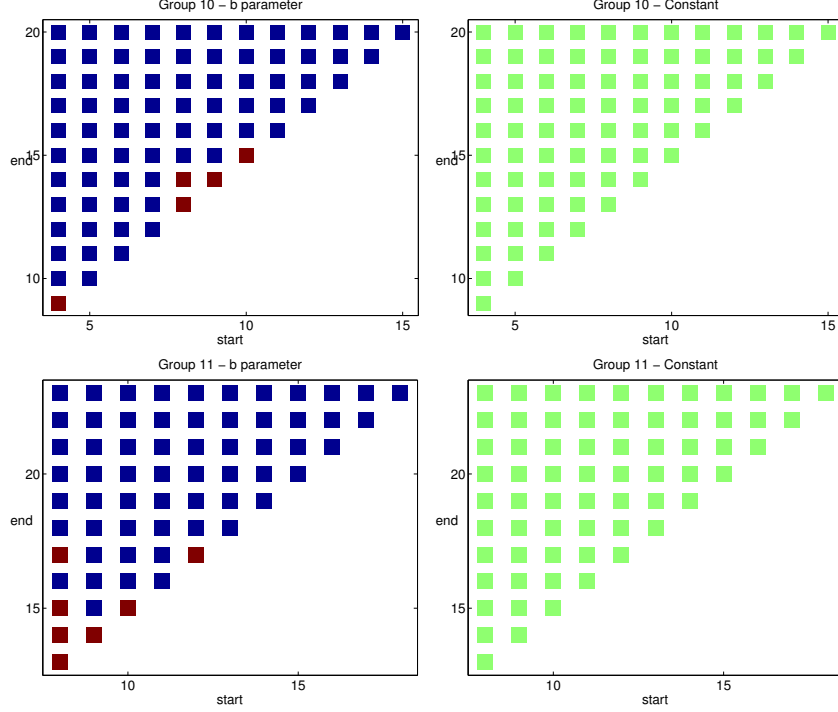
Payoff Table

The earned points are based on the following formula:

$$\text{points} = \max \left\{ 1300 \cdot \left(1 - \frac{\text{error}^2}{49} \right), 0 \right\},$$

where the error is the absolute difference between the realized and predicted price in period t .

error	points	error	points	error	points	error	points	error	points
0.1	1300	1.5	1240	2.9	1077	4.3	809	5.7	438
0.15	1299	1.55	1236	2.95	1069	4.35	798	5.75	423
0.2	1299	1.6	1232	3	1061	4.4	786	5.8	408
0.25	1298	1.65	1228	3.05	1053	4.45	775	5.85	392
0.3	1298	1.7	1223	3.1	1045	4.5	763	5.9	376
0.35	1297	1.75	1219	3.15	1037	4.55	751	5.95	361
0.4	1296	1.8	1214	3.2	1028	4.6	739	6	345
0.45	1295	1.85	1209	3.25	1020	4.65	726	6.05	329
0.5	1293	1.9	1204	3.3	1011	4.7	714	6.1	313
0.55	1292	1.95	1199	3.35	1002	4.75	701	6.15	297
0.6	1290	2	1194	3.4	993	4.8	689	6.2	280
0.65	1289	2.05	1189	3.45	984	4.85	676	6.25	264
0.7	1287	2.1	1183	3.5	975	4.9	663	6.3	247
0.75	1285	2.15	1177	3.55	966	4.95	650	6.35	230
0.8	1283	2.2	1172	3.6	956	5	637	6.4	213
0.85	1281	2.25	1166	3.65	947	5.05	623	6.45	196
0.9	1279	2.3	1160	3.7	937	5.1	610	6.5	179
0.95	1276	2.35	1153	3.75	927	5.15	596	6.55	162
1	1273	2.4	1147	3.8	917	5.2	583	6.6	144
1.05	1271	2.45	1141	3.85	907	5.25	569	6.65	127
1.1	1268	2.5	1134	3.9	896	5.3	555	6.7	109
1.15	1265	2.55	1127	3.95	886	5.35	541	6.75	91
1.2	1262	2.6	1121	4	876	5.4	526	6.8	73
1.25	1259	2.65	1114	4.05	865	5.45	512	6.85	55
1.3	1255	2.7	1107	4.1	854	5.5	497	6.9	37
1.35	1252	2.75	1099	4.15	843	5.55	483	6.95	19
1.4	1248	2.8	1092	4.2	832	5.6	468	error ≥ 7	0
1.45	1244	2.85	1085	4.25	821	5.65	453		



Notes: The regressions are estimated with different starting (horizontal axis) and ending (vertical axis) periods. The panels color code the significance of the estimated parameters: red - significantly positive, blue - significantly negative, green - not significantly different from 0. The left panel shows the b parameter, the right panel the constant, a .

Figure 11: Regression result of estimating Eq. (8) on small groups

Appendix B Supplementary analysis, figures and tables

B.1 Bubble-growth

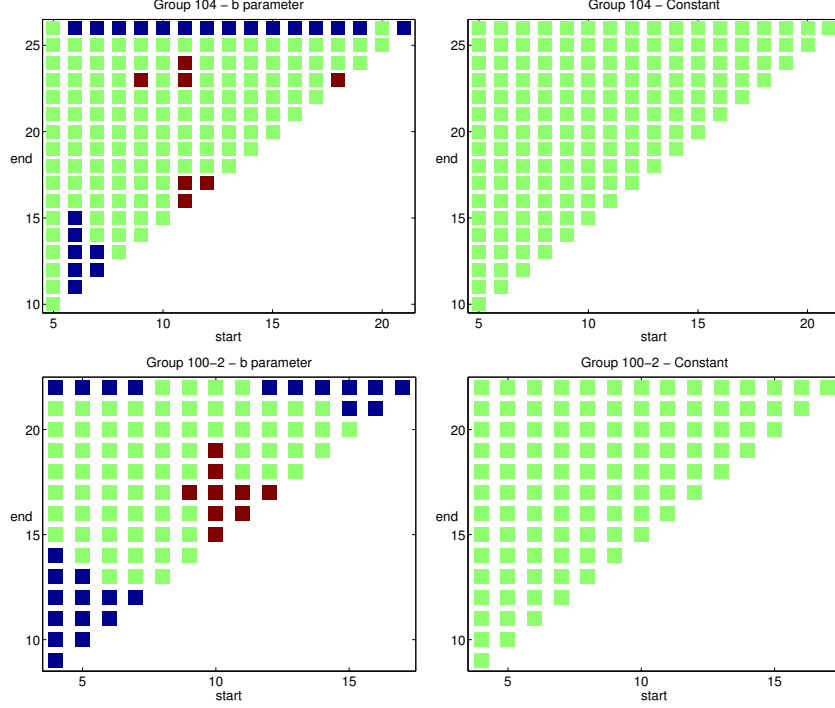
Looking at Figure 1 the price seems to increase with a higher than exponential growth rate in the large bubbles. Following Hüsler et al. (2013) we estimate the growth rate with two specifications. The first one assumes anchoring on the price, and uses the following equation:

$$\log \left(\frac{\bar{p}_t}{\bar{p}_{t-1}} \right) = a_1 + b_1 \bar{p}_{t-1} \quad (8)$$

The second specification is based on anchoring on return:

$$\log \left(\frac{\bar{p}_{t+1}}{\bar{p}_t} \right) = a_2 + b_2 \log \left(\frac{\bar{p}_t}{\bar{p}_{t-1}} \right) \quad (9)$$

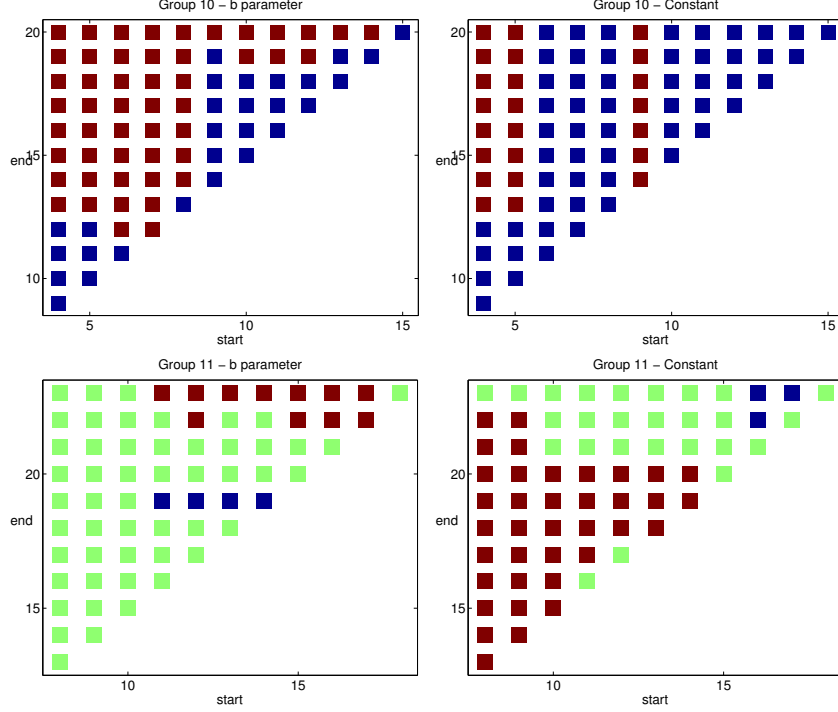
In both cases, if $a_i > 0$ and $b_i > 0$ ($b_i < 0$), then the growth is larger (smaller) than exponential, but the feedback is based on prices in the first, and on return in the second case. Looking at Figure 1 again, we can see two large and two small markets with very large bubbles. These are markets 10 and 11 for the small groups, and markets 100-2 and 104 for the large groups. The bubble periods are different for the



Notes: See explanation under Figure 11.

Figure 12: Regression result of estimating Eq. (8) on large groups

different groups. The starting period is when the price exceeds the fundamental for the first time, and the ending period is at the price peak. We estimated the parameters with different starting and ending periods, as it might be that the growth rate is different at the beginning of the bubble than towards the end. Figures 11-14 display the significance of the coefficients of the regression results: blue means significantly negative, red means significantly positive, and green means insignificant coefficient. The actual parameters are between -0.47 (min at a_2 for Group 104) and 3.19 (max at b_2 for Group 104) in all cases. Considering a price anchor, we find that in the small group the price increases with a lower than exponential rate for almost all specifications, whereas the growth rate is faster in the large group (but still not significantly faster than exponential growth in most specifications). This can also be seen on Figure 1. However, if we look at Figures 13 and 14 there is not so clear difference between group sizes. Anchoring on return results in a faster than exponential growth in both markets for early starting rounds, whereas a slower than exponential growth rate towards the peak. Also, in this case parameter a_2 is significantly positive for early starting points. These results give an indication that growth rate might be higher in the large groups than in the small groups, but we cannot draw strong conclusion given the low number of observations.



Notes: See explanation under Figure 11.

Figure 13: Regression result of estimating Eq. (9) on small groups

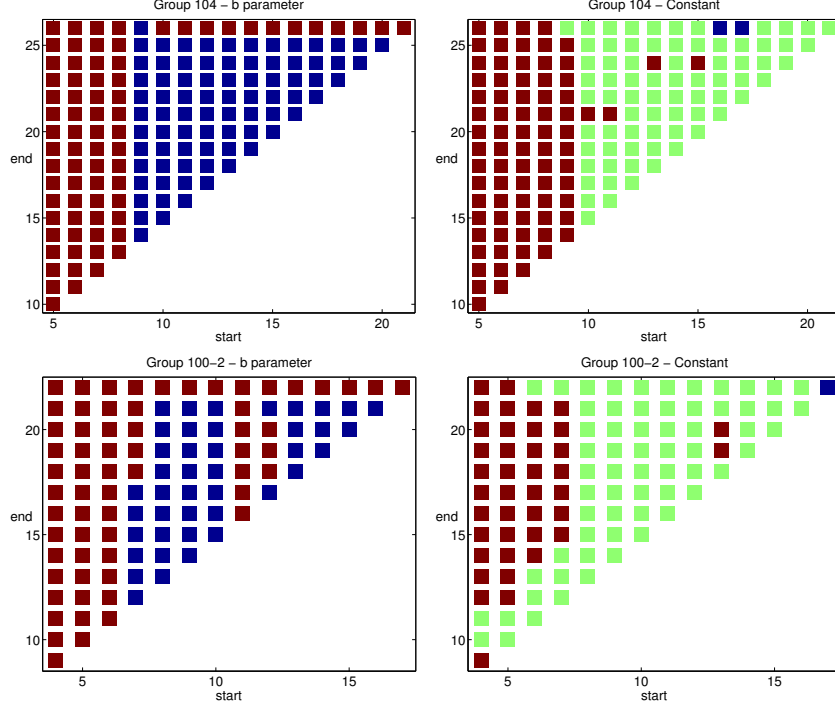
B.2 Individual behaviour

Figures 15 and 16 show the individual forecasts and the market price in each market.

B.3 Heuristic Switching Model

Figures 17-19 show the HSM simulated for the three different behavioural pattern we have observed in the experiment. The upper panel on each figure shows the experimental price and the simulated price for the given market. As we can see, in all of the cases the simulation followed the experimental pattern very well. In the left bottom panel we can see the forecasts of the different rules (left panel) and the corresponding forecast errors (right panel). For stable markets, the forecasts are relatively close to each other, with very small forecast error. As we move on to more unstable markets, we can see that forecasts are more heterogeneous, and forecast errors increase. On the right panel the evolution of the fractions are shown. In the stable markets, there is no clear pattern which rule dominates the market, as rules more or less yield to the same payoff.

As we can seen in these figures, the HSM does a good job in describing the experimental pattern. To quantify the model performance, we look at different benchmark models with and without heterogeneity,



Notes: See explanation under Figure 11.

Figure 14: Regression result of estimating Eq. (9) on large groups

and determine the mean-squared error (MSE) of these models by calculating the average squared difference between the simulated and the experimental price. We considered 6 different homogeneous rules: the four rules we have used in the HSM, plus the fundamental rule ($p_{t+1}^e = p^f$ for all t) and the naïve rule ($p_{t+1}^e = p_{t-1}$). Furthermore, we looked at the heterogeneous population using the four rules of the HSM each with equal weight, and the benchmark HSM rule we used in the previous section. Finally, we have also performed a grid search in steps of 0.01 for $\beta \in [0, 10]$, $\eta \in [0, 1]$ and $\delta \in [0, 1]$ in order to find the best-performing HSM-model for each market.

Table 5 lists the MSE for each market for each model, and the parameters for the best-fitting HSM-model.²⁶ The markets are ordered in the same way as in Table 1. If we look at the homogeneous rules, we can see that for more stable markets the WTR and LAA are the best performers, whereas the STR captures the more unstable markets better. For the more stable markets, there is no substantial differences between the homogenous rules (excluding the fundamental rule). Naturally the magnitude of MSE is much smaller for these markets than for the more unstable one. There the variance in performance of the different rules are much larger as well. In general we can see that subjects do not follow the fundamental rule. It produces a

²⁶In most of the markets, the best parameter set is unique. There is a multiplicity of parameter sets in Group 6 (3 sets), Group 100-1 (14), Group 100-2 (108), and Group 140 (1672). In this case we report the set with the lowest values.

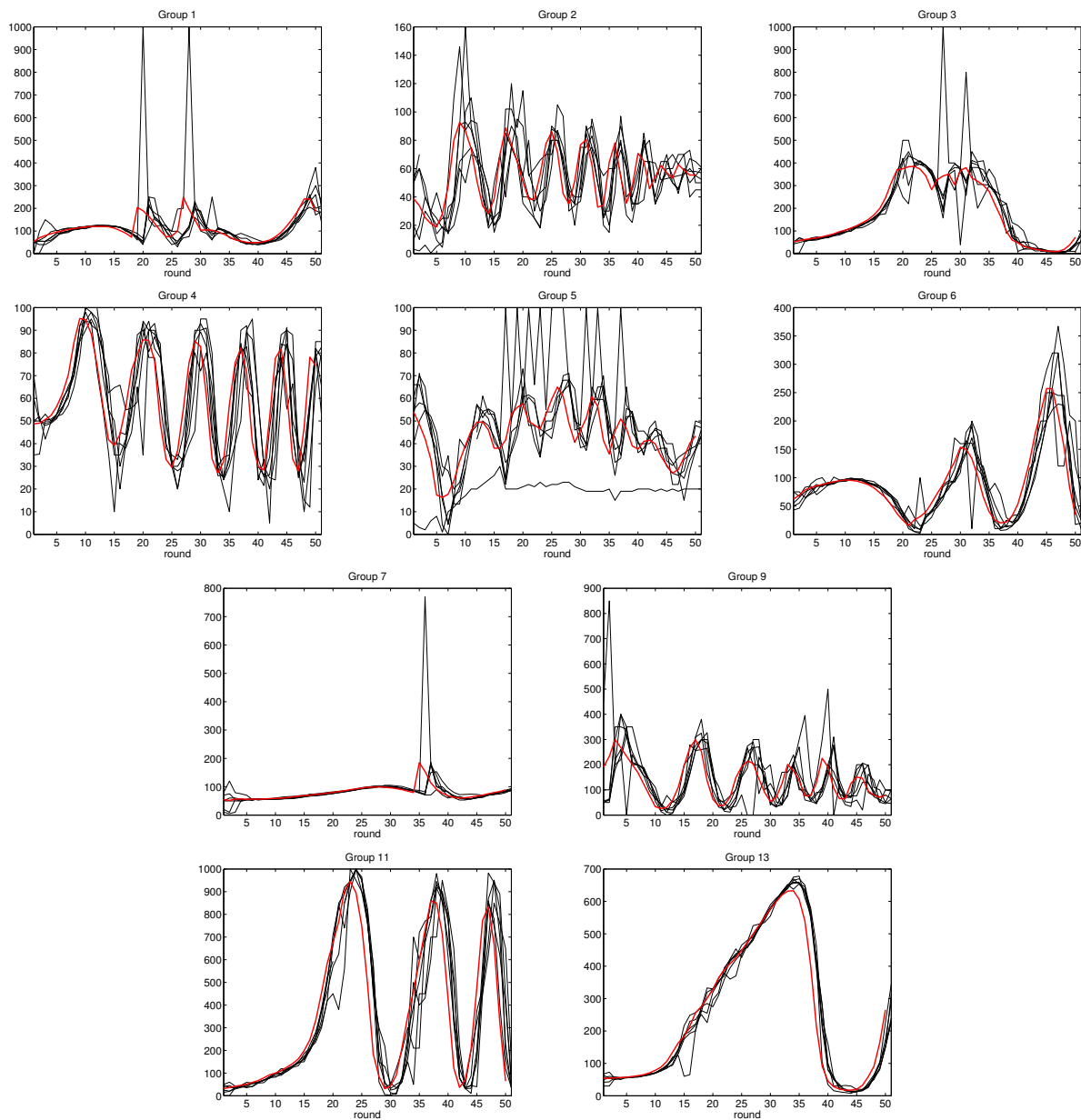


Figure 15: Further small markets

quite high MSE in each market. If we turn to the heterogeneous rules, we can see that in most cases the benchmark HSM performs the best (after the fitted model which is the best by definition). There are two markets in which the fixed fraction performs better. In these cases it seems that the benchmark HSM has parameters that are quite different from the optimal one, which can cause a decline in the performance of the benchmark HSM. In most of the markets, allowing for heterogeneity, and further flexibility with switching substantially improves fit. Subjects do not seem to stick to one decision rule if it proves to be inefficient.

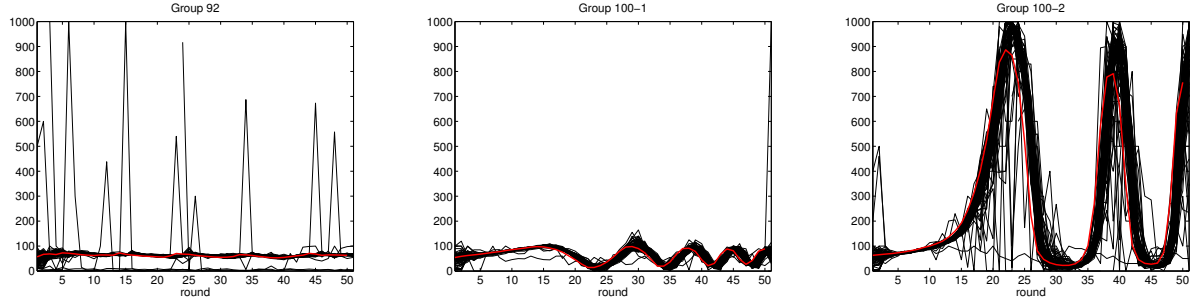


Figure 16: Further large markets

	Homogeneous rules						Heterogeneous rules			Fitted parameters		
	Fund	Naïve	ADA	WTR	STR	LAA	Fixed fraction	Original HSM	Fitted HSM	β	η	δ
Gr. 5	673.66	39.53	57.43	30.15	52.62	33.57	25.23	24.98	24.4	10	0.69	0.95
Gr. 2	407.72	255.82	340.32	200.5	352.36	118.71	154.3	98.43	93.71	0.21	0.74	0.82
Gr. 4	469.58	230.69	351.9	152.37	224.14	63.4	103.5	58.25	53.32	0.52	0.35	0.87
Gr. 8	50.31	1.89	2.83	1.58	3.04	5.5	1.84	1.67	1.59	10	0.01	0.84
Gr. 7	799.57	307.23	320.47	363.81	833.58	500.2	381.39	638.06	319.07	0.01	0.25	0.14
Gr. 6	4192.6	603.64	1159.3	284.58	155.96	546.5	205.84	159.39	87.41	1.33	1	0.97
Gr. 12	17542	2930.8	5171.8	1617	1460	1887.8	1140.3	645.82	639.36	0.12	0.54	0.92
Gr. 1	5332.8	1283.5	1494.7	1357.7	2859.7	1615.6	1349.4	1631.3	1255.45	10	0	0.67
Gr. 9	8561.3	2252.5	3514.1	1473.6	2124.3	882.85	1038.9	692.02	625.41	0.02	0.5	0.77
Gr. 3	32848	963.03	1571.2	729.01	1164.6	5119.1	1141.2	1004.5	592.21	10	0.17	0.65
Gr. 13	83711	2479	4774.1	1239.2	364.78	11390	1998.4	455.97	364.78	1.35	0.8	0
Gr. 11	189950	20763	36893	10940	3521.4	16092	8110.2	2116.4	1657.85	0.08	0.9	0.87
Gr. 10	268890	16112	30978	7915.4	2430.5	18773	6880.8	1552	1456.28	6.16	0.18	0.86
Gr. 92	40.25	11.44	14.63	11.34	24.62	14.51	11.24	11.14	10.77	10	0.85	0.98
Gr. 103	21.08	4.1	7.17	2.45	3.57	2.45	1.71	1.68	1.21	5.36	0.11	0.78
Gr. 100-1	730.19	274.43	431.68	172.21	216.93	107.44	115.3	53.32	43.29	9.59	0.54	0.8
Gr. 99	7112.4	1695.2	2770	983.26	862.51	664.35	624.24	258.11	244.25	0.02	0.9	0.87
Gr. 100-2	132080	17242	31029	9289.1	4929.1	13738	7223.1	2974.5	2911.29	4.78	0.76	0.92
Gr. 104	124200	12132	22263	6476.6	2570	14661	5469.9	1661.4	1099.34	1.65	0.87	0.38

Notes: Table contains the mean squared error for the different rules for each market. Markets are ordered as in Table 1. For fixed fractions the last four rules are used with 25% weight each. The original HSM contains the models simulated in Section 5.2 with $\beta = 0.4$, $\eta = 0.7$, and $\delta = 0.9$. The fitted HSM contains MSE of the grid search. The corresponding parameters are presented in the last three columns. Numbers on bold represent the lowest MSE for homogenous and heterogenous rules (other than the fitted).

Table 5: Mean squared error for the different heuristics

Looking at the optimal parameters for the fitted HSM model, we can see that there is a quite big dispersion for β . There is no clear relationship between market behaviour and the best β parameter. In some markets

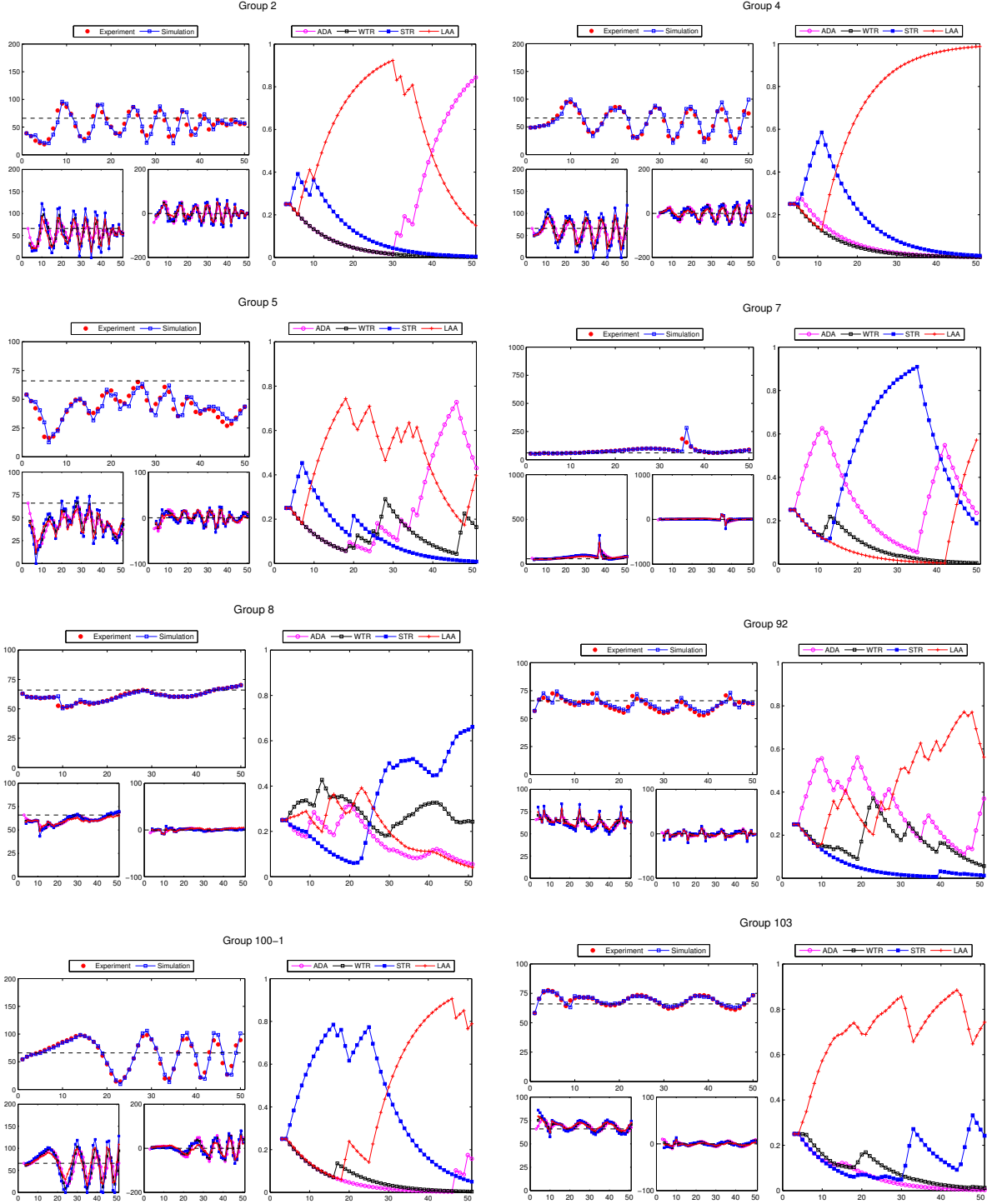


Figure 17: HSM for stable markets in a small (first 5 panels) and a large (last 3 panels) market

agents behave as if they were learning and switching to the optimal rule very quickly, in some other markets they learn in a much slower rate. Considering η and δ we can see less dispersion, but on average both parameters are slightly lower than for the benchmark HSM we have run. This suggest that our

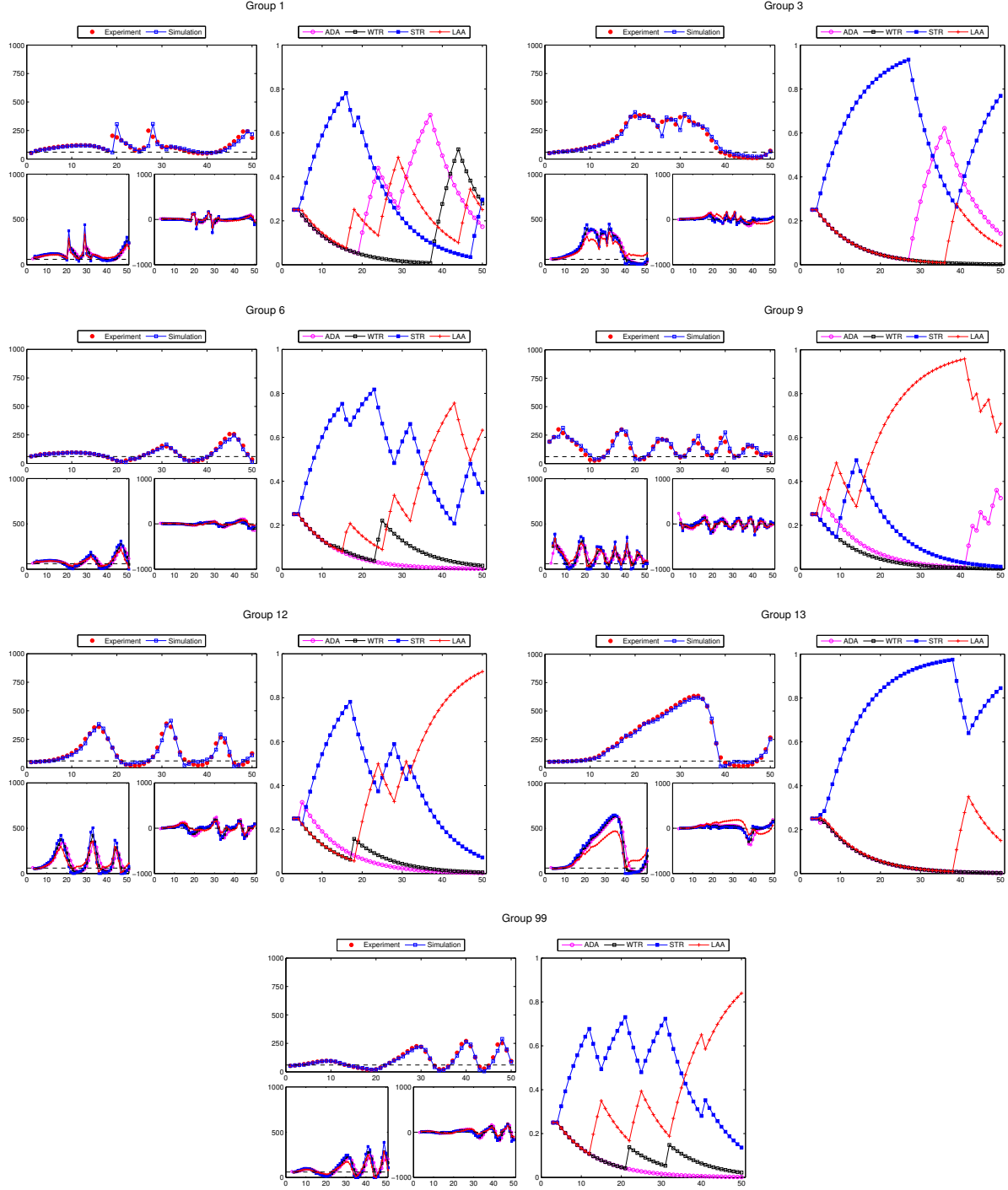


Figure 18: HSM for large bubbles in a small (first 6 panels) and a large (last panel) market

subjects consider less observations from the further past, and are willing to switch more often than in the benchmark model.

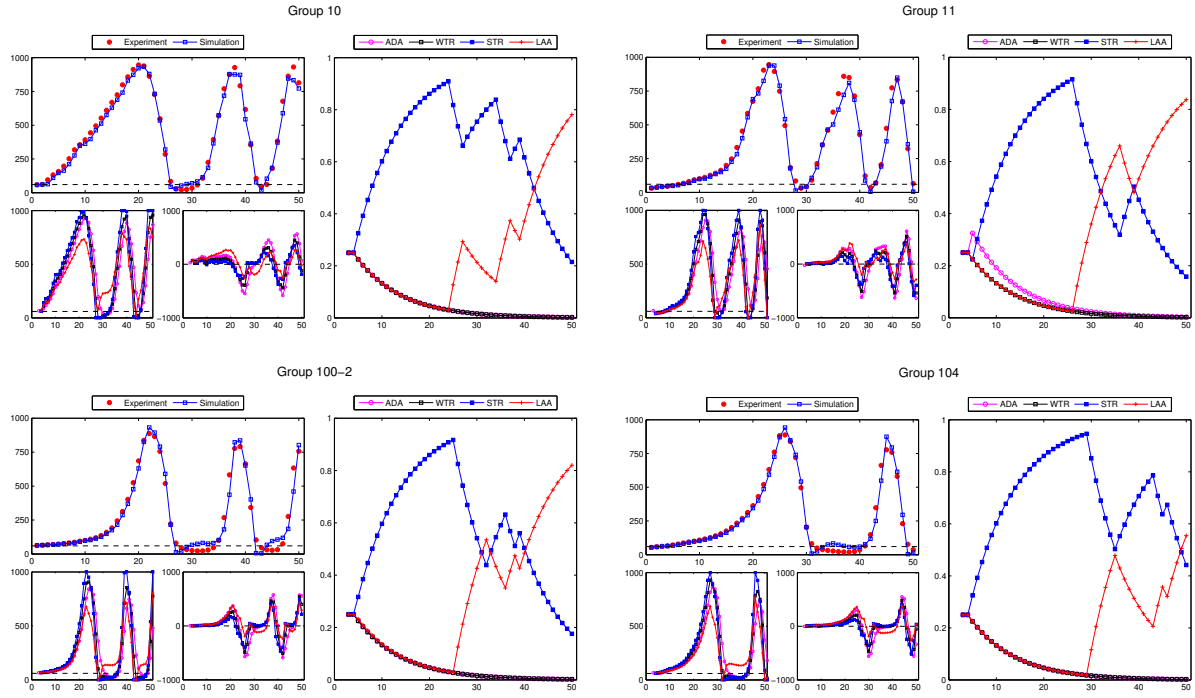


Figure 19: HSM for very large bubbles in a small (upper) and a large (lower panels) market