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

Cars Hommes[1], Anita Kopányi-Peuker[1] and Joep Sonnemans[1]

[1]University of Amsterdam and Tinbergen Institute

February 15, 2019

Abstract

We study the emergence of bubbles in a laboratory experiment with large groups of individuals. The realized price is the aggregation of the forecasts of a group of individuals, with positive expectations feedback through speculative demand. When prices deviate from fundamental value, a random selection of participants receives news about overvaluation. Our findings are: (i) large asset bubbles occur in large groups, (ii) information contagion through news affects behaviour and may break the coordination on a bubble, (iii) time varying heterogeneity provides an accurate explanation of bubble formation and crashes, and (iv) bubbles are strongly amplified by coordination on trend-extrapolation.

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

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

Acknowledgements: The authors are grateful for the support of the experimental team at the University of Valencia. We are also grateful for the discussions with the IBSEN-members, seminar participants at the University of Zurich and conference participants at ABEE 2016, EF 2016, CEF 2016, CCS 2016, ESA 2016, EEA-ESEM 2018, BEAM-ABEE 2018. 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).

Address: Amsterdam School of Economics, UvA, Roetterstraat 11, 1018WB Amsterdam, Netherlands

E-mail: C.H.Hommes@uva.nl; A.G.Kopanyi-Peuker@uva.nl; joepsonnemans@gmail.com

1 Introduction

Expectations play an important role for intertemporal decisions in everyday economic life. Based on their expectations about the future, agents decide on what actions to take today. For example, traders and firms form price forecasts and based on these forecasts they buy or sell assets or decide on how many goods to produce. The subsequent actions that all individual agents then execute determine the realised market price and aggregate behaviour. To understand markets we have to understand how groups of individuals form and coordinate their expectations. Since Muth (1961) and Lucas (1972) the traditional approach has become the Rational Expectations (RE) hypothesis, which states that the expectations of all agents are the same and consistent with their model of the economy. While the RE hypothesis may be a natural benchmark, as a realistic description of real world behavior it faces challenges both theoretically and empirically. For example, many studies show that survey data on expectations are inconsistent with RE, see e.g. the recent survey of Coibion et al. (2018). Alternative models of expectations assume that agents are *boundedly rational* (Sargent, 1993), are prone to behavioral biases (Barberis and Thaler, 2003), form expectations through an adaptive learning process (Evans and Honkapohja, 2001; Branch and Evans, 2010) or use simple, but 'smart' heuristics (Anufriev et al., 2018).

Speculative asset markets are probably the best example where price movements are amplified by expectations ('animal spirits') leading to longlasting bubbles followed by sudden market crashes. Recent examples of bubbles include the dot com stock market bubble in the late 1990s, the U.S. housing market bubble in the early 2000s and the recent bitcoin bubble in 2017 (and crash in 2018). It has been argued that non-rational, trend-extrapolating expectations, have amplified these movements in asset prices. For example, Case et al. (2012) study surveys of household expectations about changes in home values and show the prevalence of trend-extrapolation, while Barberis et al. (2018) argue that trend-extrapolation explains bubbles in the stock market, the housing market and commodity markets.

A complementary method to simultaneously study expectation formation of a group of individuals and the emergence of bubbles is a controlled laboratory experiment. Bubbles have been extensively studied in the lab, for example in the seminal work of Smith et al. (1988) and follow-up papers (see the extensive overview of Noussair and Tucker, 2013). Almost all of these bubble experiments, however, use small groups of 6-10 subjects. The main goal of the current paper is to study whether bubble formation is robust to increasing the group size. Will a large group of individuals coordinate expectations on trend-extrapolating rules and cause an asset market bubble? To address this question, we conduct asset market experiments with large groups of up to 100 subjects, to our best knowledge the largest groups of paid human subjects in bubble experiments[1].

Our experimental design is based on the learning-to-forecast asset pricing experiment of Hommes et al. (2005, 2008)[2]. Subjects need to forecast the future price of a risky asset for 50 periods. The average forecast of the group will determine the realised market price with a positive feedback between expectations and realisations: higher (lower) expected prices lead to higher (lower) market prices[3]. Subjects know this qualitative relationship, but they do not know the exact law of motion for the prices. The market price is determined as an equilibrium price using mean-variance optimisation of asset trade, which is computerised, and subjects do not need to carry out the trades themselves. Subjects are paid according to their forecasting performance. In Hommes et al. (2008) very large bubbles in 5 out of the 6 groups have been observed. In their setup the market price depended on the average forecast of a group of six individuals, so that the influence of one subject was relatively high. Bubbles may then arise because of one "irrational" individual. The natural question is whether bubbles would still arise in large markets, where each individual only has a small weight in determining the market price. In our experiment about 100 subjects form a large market. 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, ranging from 92 to 104 subjects and 13 small groups of 6. Such large groups do not fit in a single laboratory and we coupled the CREED lab at UvA and the LINEEX lab in Valencia. A large group thus consists of about 50 subjects in Amsterdam and 50 in Valencia and their overall average price forecast determines the realized market price[4].

[1]Duffy (2016) gives an overview of the progress in the related field of experimental macroeconomics. A common critique to these macroeconomic experiments is the limited group size. In the real macro economy many agents are interacting with each other and, therefore the findings in small group experiments might not be directly applicable outside the lab. Our study contributes to the important issue of how group size may affect individual and aggregate behavior.

[2]Learning-to-Forecast Experiments (LtFE) were pioneered by Marimon and Sunder (1993); Marimon et al. (1993); 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 all actions based on the expectations are computerised based on optimal decision making. In this way clean data on expectations are obtained, as subjects are only rewarded based on their forecasting performance. A large literature on LtFEs in different economic environments has developed, see Hommes (2011) for an overview.

[3]Earlier work has shown 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 experimental evidence see Heemeijer et al. (2009), Sonnemans and Tuinstra (2010) and Hanaki et al. (2019).

[4]LtFEs are related to repeated number guessing games or beauty contest games, introduced by Nagel (1995). In number guessing games subjects predict a number between 0 and 100 and the winner is she whose guess is closest to 2/3 of the average. The Nash equilibrium of this guessing game is 0, but in the laboratory experiment first- and second order rationality (where the subject guesses $2/3 \cdot 50$ respectively $(2/3)^2 \cdot 50$) are most common. These games differ is several aspects from LtFE: (i) In number guessing games the Nash Equilibrium is at the border (so undershooting is not possible), while in LtFE the equilibrium price is strictly positive; (ii) number guessing games are competitive and only the subject who's guess is closest to the target (the winner) earns a positive payoff, while in LtFE payoff is determined by the individual size of the error in

It is important to note that a priori theory can not decide whether large groups are more or less stable than small groups. One could argue that large markets are more stable, as each individual has a smaller effect on the market, and individual errors are more likely to cancel out (Muth, 1961). On the other hand, once a bubble arises in a large group, e.g. due to a few positive shocks and coordination of a majority on trend-following behaviour, it may be hard to break the coordination on a bubble in a large group. This suggests that large groups may be more unstable. Hence, theory does not provide a unique answer, and it becomes an empirical question, well suited for a laboratory test, whether large markets are more or less stable.

Beyond the large group size and the coupling of two different labs, we introduce another new experimental design feature in LtF experiments, namely the role of "news" elements and its effect on individual and aggregate behavior. 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. We introduce news announcements when prices are too high or too low compared to the fundamental value. These news elements are short messages that can appear on a subjects' screen: "Experts say the stock market is overvalued (undervalued)." It is common knowledge that these 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.[5] A key question then is whether *information contagion* arises, that is, whether news about fundamentals can break the coordination on a bubble in a large group. In order to prevent prices diverging forever, we also kept an (unknown) upper bound on the forecasts, but it is high enough to leave room for the news to have an effect before prices reach the upper bound.

Our main results are fourfold. First, bubbles are robust in large groups: two out of six large groups exhibit very large bubbles of more than ten times fundamental until they reach the exogenously set upper bound. Three other large groups are rather stable with small fluctuations around the fundamental. Second, we find information contagion at the individual and the aggregate level. At the individual level subjects who receive news about an increasing asset bubble have a significantly lower increase in their forecasts

the prediction, (iii) the feedback parameter in number guessing games (e.g. 2/3) is typically much farther away from 1 than in LtFE (in this experiment 0.95). See Sonnemans and Tuinstra (2010) for an overview and an experimental comparison of these two games. See Mauersberger and Nagel (2018) for an extensive overview of experimental coordination games and their relevance for macro and finance.

[5] In a different context (monetary policy in a macroeconomic liquidity trap experiment) Hommes et al. (2018) provide all participants with (bad) news in specific pre-determined periods. In our setup news announcements arise endogenously depending only on the price and only for a randomly chosen subset of participants.

compared to those who did not receive the news. At the aggregate level news about an increasing bubble has an effect on the price in some (six out of eight) of the small groups and in one (out of three) large groups. In these groups the news breaks the coordination on the bubble before reaching the upper-bound and leads to a market crash. Third, analysis of the individual and aggregate data of the large groups provides an accurate explanation of the bubbles and crashes. At the initial stage of the experiment expectations heterogeneity is high, but quickly subjects coordinate their expectations, often on a non fundamental price. Due to small random shocks and coordination on a trend-extrapolating forecasting rule the asset price starts increasing slowly. Strong coordination of expectations amplifies the bubble. As the bubble forms, expectations heterogeneity gradually increases and eventually the market crashes. Heterogeneity then peaks and remains high suggesting a state of confusion after the market crash. This pattern clearly emerges for the large bubbles in the large groups. The same pattern seems to occur in small groups, but it is noisier. Fourth, we calibrate the behavioural heuristics switching model (HSM) of Anufriev and Hommes (2012) to both small and large groups. Agents switch between simple forecasting heuristics based upon their relative performance. For the unstable markets, both for small and large groups, in the first 30 periods agents coordinate expectations on a strong trend-extrapolating rule. The HSM thus provides a behavioral explanation of the amplification of the bubble by trend-extrapolating expectations for both group sizes.

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 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. A related paper to our study is Bao et al. (2016), who investigated a similar setup in 7 groups, with size ranging from 21 to 32 subjects. In their experiment 6 out of the 7 groups exhibit large bubbles, and these bubbles arise faster than in the small groups of Hommes et al. (2008). We add to their work by larger group size (a size that does not fit a single lab), thus obtaining a more accurate picture of individual and aggregate behavior in bubble formation, and by studying information contagion of news about fundamentals in small and large groups.

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. In section 4 we describe the behavioural heuristics switching model, and its fit on our experimental data. Section 5 concludes.

2 Experimental design

2.1 Experimental market

The lab environment is based on an asset pricing model with heterogeneous beliefs following Brock and Hommes (1998); for a textbook treatment see Campbell et al. (1997)). 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 $\bar{y}$. Based on this allocation decision, the wealth of agent i in period $t+1$ is given by

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

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:

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

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 defined as $\rho_{t+1} \equiv p_{t+1} + y_{t+1} - Rp_t$. 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

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

where $p_{i,t+1}^e = E_{i,t}(p_{t+1})$ is the individual forecast by agent i of the price in period $t+1$ and $\bar{y} = E_t[y_t]$ is the forecast of the exogenous dividend process y_t, assumed to be correct for all agents. The market price for the risky asset is set by market clearing, that is demand equals supply:

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

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 results in an equilibrium pricing equation

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

or equivalently

$$p_t = p^f + \frac{1}{1+r}(\overline{p}^e_{t+1} - p^f) + \varepsilon_t, \tag{5}$$

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

Note that the realized price p_t in period t depends on the average prediction $\bar{p}^e_{t+1}$ for period $t+1$ and agents use information up to period $t-1$ when predicting the price of $t+1$, so all forecasts are two periods ahead. 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 becomes $p_t = p^f + \varepsilon_t$, so that expectations are self-fulfilling. In the experiment, we use the following parameters: $\bar{y} = 3.3$ and $R = 1.05$. This implies that the fundamental price is 66.

Notice also that the price equation(5) has rational bubble solutions, where the deviation from the fundamental price $x_t = p_t - p^f$ grows at the risk-free rate r. In theory, these rational bubbles are often excluded by a transversality condition.

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 positive feedback in the market (i.e., the higher their forecast is, the higher the realized price will be ceteris paribus).

Subjects are only paid according to their forecasting performance. Their payoff is determined by the following formula:[6]

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

where $p^e_{i,t+1}$ denotes the forecast of the 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 payoff function, and they were also provided a payoff table showing earnings corresponding to given forecast

[6] The payoff function is the same as in Hommes et al. (2005, 2008) with groups of six. 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 with groups of six subjects.

errors. Subjects accumulated their payoffs during the experiment, which was converted to euros at 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 since $\bar{y}$ and r were common knowledge 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. They would receive an error message in that case and had to reenter a forecast ≤ 1000[7]. As a new experimental design feature, 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. Whenever 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 7$^{\text{th}}$ session subjects were grouped in markets of 6. This results in total in 6 large markets and 13 small markets.[8] 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 time span, 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

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

[8] In the session for the small markets, we had 78 subjects. Additionally we have run 2 pilot sessions in Amsterdam. As here we did not connect the two labs, the procedures of the experimental sessions and pilot sessions differ. Therefore we do not consider these pilot data in the analysis.

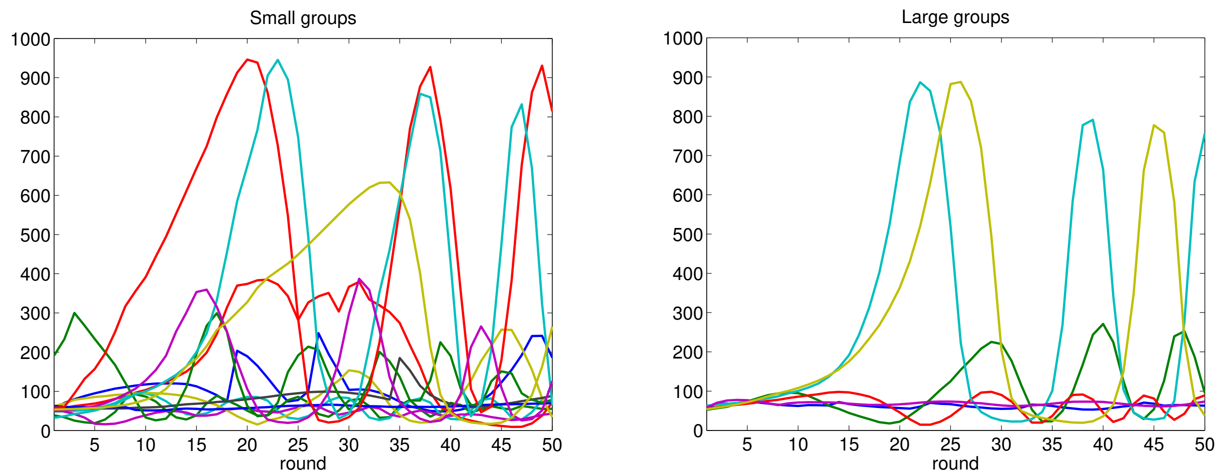


Figure 1: Realized prices in each market for the Small (left panel) and the Large (right panel) treatments

later periods. If subjects did not submit a decision on time, they would not earn anything that period, and the average forecast was determined based only 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 upon 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).[9]

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. 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).[10] The time evolution of the market price is very heterogeneous across groups for both the Small and the Large treatments. We distinguish three different qualitative market behaviours:

(i) markets which are stable or exhibit small oscillations around the fundamental;

(ii) markets with moderately large bubbles, with a peak at about 3-4 times the fundamental price, so that some subjects receive news, but the bubble does not reach the exogenous upperbound; and

(iii) markets with very large bubbles, with a peak of more than 10 times fundamental value, and a crash because some 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. The markets are ranked according to their median price. For the small group size there are five stable markets (the first five groups in Table 1), six markets with moderately large bubbles (groups 6 to 13 in Table 1) and two markets with very large bubbles (groups 10 and 11). For the large group size there are three stable markets (groups 92, 103 and 100-1), one with a moderately large bubble (group 99) and two with very large bubbles (groups 100-2 and 104). Markets which have a relative low average/mean price also have a relative low standard deviation and are thus more stable. Other markets are heavily overpriced, and they also have rather high standard deviation due to the observed bubbles and crashes.

The overall 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, $p = 0.93$).[11] 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 ($p = 0.17$; n=6).[12] All markets' standard deviations are significantly higher than the

[9] The total expenses (including payment for the volunteer's dilemma, and laboratory expenses) were about 18K euros.

[10] For one of the large groups, Group 104, the market price was incorrectly saved, leading to small errors in round 30, for 22 of the 104 subjects, who saw a price of 202.1 instead of the realized market price of 204.1. Note that this difference is small in absolute terms (about 1%) as well as compared to the distance from the fundamental price 66. For the data analysis we used the market price of 204.1 in that round for all subjects.

[11] All statistical tests are at the group level ($n = 13$ for the small groups and $n = 6$ for the large groups), using nonparametric tests.

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

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 group 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. Note that $IE = DE + CE$.

Table 1: Descriptive statistics of the markets

standard deviation predicted by the rational expectation model, that is, the $SD = 0.5$ of the noise term ε_t in (5), so all markets exhibit significant excess volatility.

Table 1 also reports the Relative Absolute Deviation (RAD) and the Relative Deviation (RD) from the fundamental price.[13] Most markets are overpriced, but 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:

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

group 5 in the Small treatment (group 8 shows almost the same pattern). The most stable groups 92, 103 and 8 (in terms of SD) 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, based on the market price and these summary statistics, no striking differences between the aggregate behaviour in small and large groups can be observed.[14]

3.2 Coordination of expectations

Do subjects manage to coordinate their expectations in a market or does heterogeneity prevail? Figure 2 shows time series plots of the coefficient of variation of individual forecasts, that is, the standard deviation divided by the mean, for the small and the large markets. A low (high) value of the coefficient of variation corresponds to a high (low) degree of coordination of forecasts. Heterogeneity strongly fluctuates over time with many high peaks. In the small groups, the time variation seems somewhat noisier, with many sudden high peaks perhaps due to individual experimentation. In the large groups heterogeneity seems to vary more smoothly and gradual, although there are still some sudden high peaks. In any case, we do not observe stronger coordination towards the end of the experiment, but for most markets heterogeneity continues to fluctuate over time.

To study the time-varying heterogeneity in more detail, Figures 3 to 5 plot typical examples of the three different types of market behavior for both small and large markets. These figures show the market price together with the coefficient of variation of individual predictions. Figure 3 illustrates a typical example of a stable market; Figure 4 of moderately large bubbles, and Figure 5 of very large bubbles.[15] The blue solid lines correspond to the market price and are depicted on the left vertical axis; the red dashed lines correspond to the coefficient of variation of individual forecasts and are measured on the right vertical axis.

In the stable markets (Figure 3) after an initial phase of high heterogeneity and some initial fluctuations, the heterogeneity drops almost to zero after about 15 periods. Subjects thus learn to coordinate

[14] Comparing small and large markets in Figure 1 it seems that along the very 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 suggestive evidence on this issue. This analysis is relegated to Appendix B.1.

[15] The markets have been chosen to illustrate the typical behavioural patterns. Plots for other markets and plots showing individual predictions as well as market prices per market are relegated to Appendix B.2.

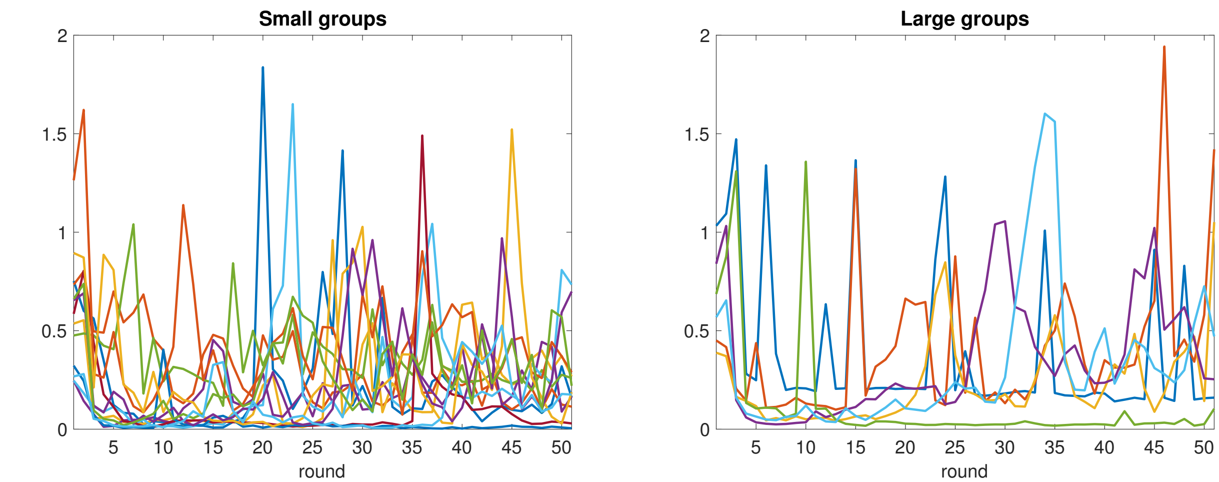


Figure 2: Coefficient of variation of individual predictions (standard deviation divided by the mean) in each market for the Small (left panel) and the Large (right panel) treatments. A low (high) value means that individual forecasts are strongly (weakly) coordinated.

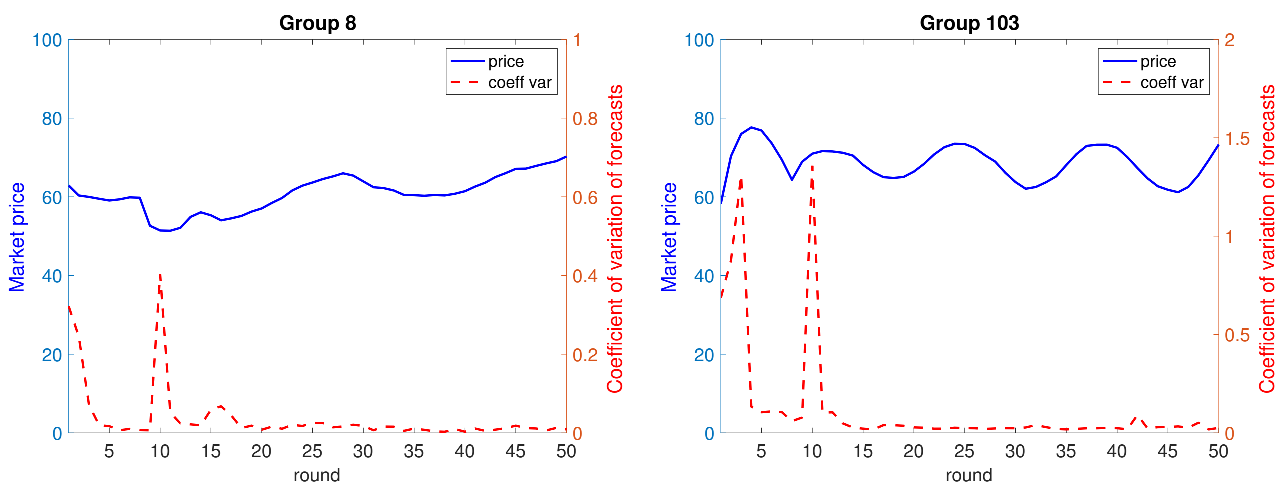


Figure 3: Market price (left scale) and coefficient of variation of individual forecasts (right scale) for examples of stable markets in a small (left) and a large (right) group.

their expectations within 15 periods and in both the small and large markets the price exhibits small fluctuations close to the fundamental value. Compared to the other markets, coordination of expectations is the strongest in the stable markets.

In the markets with moderately large bubbles (Figure 4) heterogeneity is initially high, but quickly drops to lower levels. The market price is above fundamental value, however, and starts to increase further. In the small market (left panel in Figure 4) the price starts following a strong upward trend and heterogeneity gradually increases. The market price increases and peaks around 400 in period 15, after which the market crashes and heterogeneity increases rapidly and sharply. The pattern in the large market is

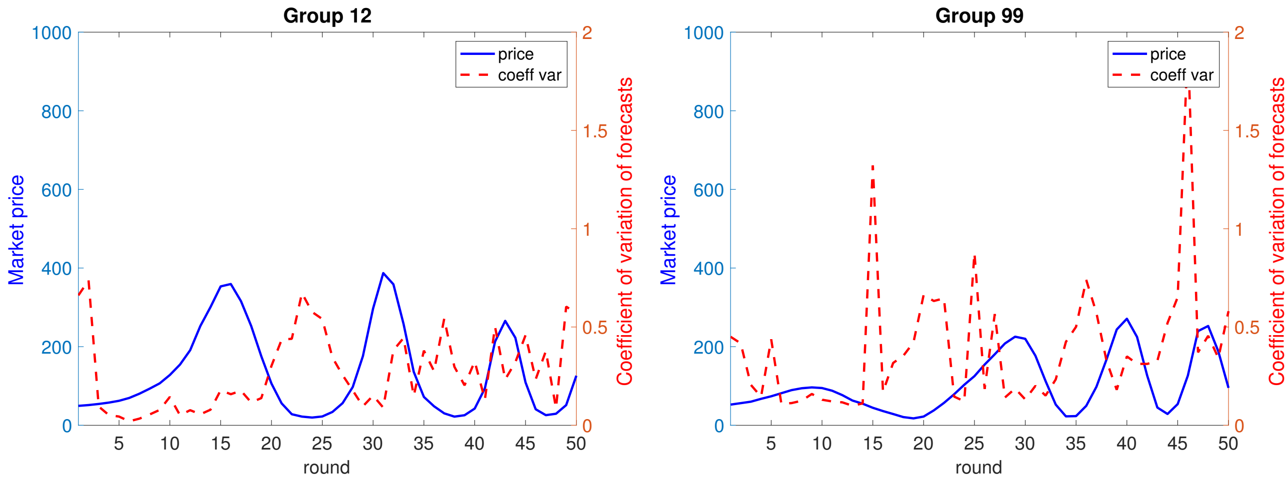


Figure 4: Market price (left scale) and coefficient of variation of individual forecasts (right scale) for examples of moderately large bubbles in a small (left) and a large (right) market.

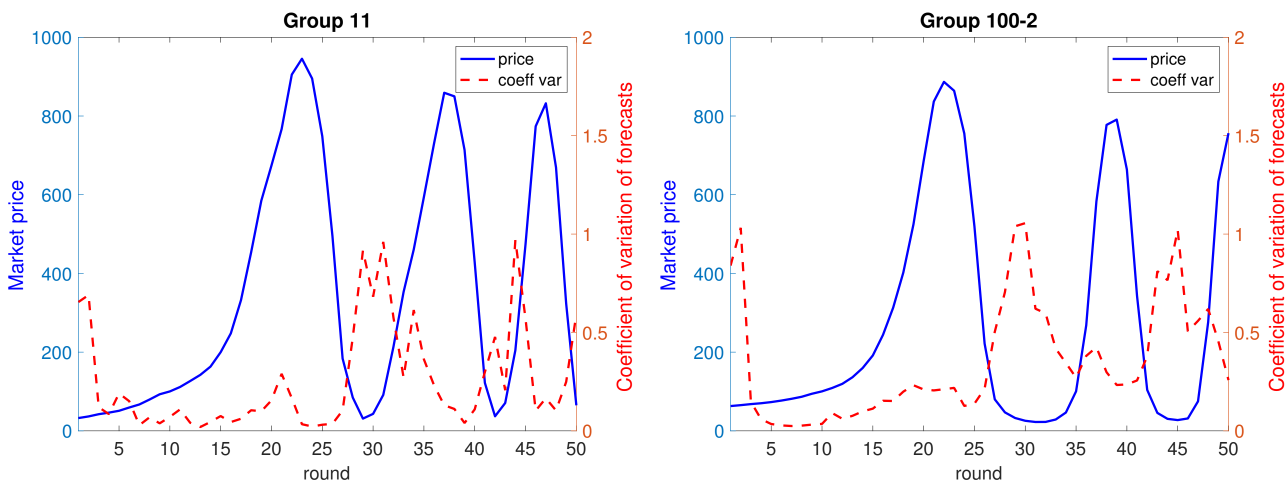


Figure 5: Market price (left scale) and coefficient of variation of individual forecasts (right scale) for examples of very large bubbles in a small (left) and a large (right) market.

somewhat different (right panel in Figure 4), where first a small bubble with a peak around 100 at period 10 occurs, followed by a decline to very low prices around period 20, followed by a larger bubble with a peak around 200 in period 30. After the first small bubble and crash, heterogeneity increases substantially and remains high during the second bubble and crash.

Figure 5 illustrates the time variation of heterogeneity in markets with very large bubbles. Initially heterogeneity is high, but quickly drops to low levels within 5 periods, after which heterogeneity gradually increases along the large bubbles. After the market crash heterogeneity peaks and remains high illustrating a state of confusion among subjects. This pattern of time variation of heterogeneity and coordination is particularly clear and smooth in the large market (right panel in Figure 5), but it also occurs in the small market, although the pattern is noisier there due to individual experimentation. We conclude from this pattern that strong coordination of expectations amplifies the large bubbles and that after a market crash heterogeneity (confusion) remains high.

To further 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 individual quadratic forecast errors for each period in which a forecast was submitted, and then taking the average of all these forecasts over individuals and periods for each market (IE $= \frac{1}{K}\sum_{t,i}(p^e_{i,t} - p_t)^2$, where K is the number of individual forecasts in a group in all 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 $= \frac{1}{K}\sum_{t,i}(p^e_{i,t} - \bar{p}^e_t)^2$, where K is the same as before). DE 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 $= \frac{1}{50}\sum_t(\bar{p}^e_t - p_t)^2$). The common error measures the quality of the average expectations. Table 1 presents IE, DE and CE for each market. By definition $IE = DE + CE$. 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 across markets whether the common or the dispersion errors are relatively large 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).

3.3 News announcements

Does news about the over- or undervaluation affect individual expectations and aggregate behaviour? In particular, can news break the coordination on a bubble? News appeared with probability 25% when the last price was sufficiently far away from the fundamental, either below 1/3 or above 3 times fundamental

	1st bubble	2nd bubble	3rd 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 as shown in the last row of 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

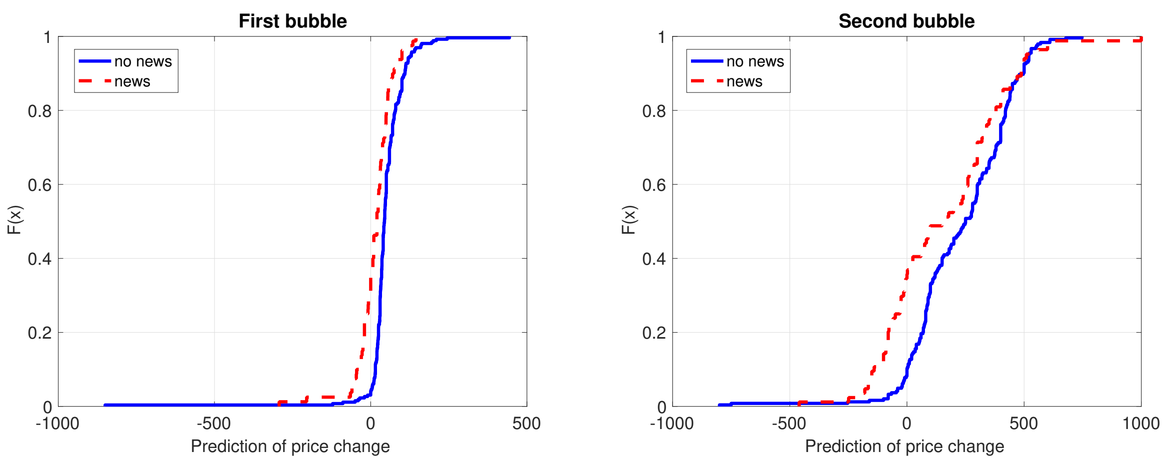


Figure 6: Cumulative distribution function of predicted price changes depending on whether the news has been seen, along the first (left figure) and second (right figure) bubbles. News of overvaluation leads to a reduction in forecasts.

value. For practical reasons, our analysis focuses only on overvaluation.[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 individual predictions 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 differences across treatments in any of these cells ($p > 0.17$). Subjects who saw the news

[16]News 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.).

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

in the first bubble increased their predictions with a significantly lower amount than those who have not seen the news ($p = 0.075$). Hence, news of overvaluation along the first bubble makes subjects significantly reduce their estimates. A similar reduction after news of overvaluation is seen for the second and third bubbles, but the differences are not statistically significant ($p = 0.16$ resp. $p = 0.13$). 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 of overvaluation. Another striking feature of Table 2 is that the average increase in predictions substantially grows from the first to the second bubble. 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.

News of overvaluation thus has a significant effect on individual expectations: it leads to significantly lower increase in forecasts along the first bubble. Does this effect of news on individual behaviour translate into an aggregate effect and stabilize the bubble? We do not have enough group observations to address this question statistically, but we can discuss it at least based on our qualitative evidence and, moreover, we will compare our small group treatment with the small group experiment without news in Hommes et al. (2008).

Out of the 13 small groups in our experiment, 8 receive news about overvaluation. Out of these 8 groups, only two (groups 10 and 11) exhibit very large bubbles with the price reaching the exogenous upperbound, while the other six groups (groups 6 to 13 in Table 1) exhibit only moderately large bubbles. For the small groups, the news about overvaluation thus breaks the bubble in six out of eight markets. Out of the 6 large groups, 3 receive news about overvaluation. Two of them exhibit very large bubbles with peaks of 900+, close to the exogenous upperbound, while in only one out of three cases the news about overvaluation breaks the coordination of the bubble. We may conclude that for both small and large groups news about overvaluation *can* break the coordination on a bubble.

In the 6 small groups of Hommes et al. (2008) without news announcements, all 6 groups reached a price level where news would have been provided in the current setup. Of these 6 groups, 5 groups evolve into a very large bubble (the remaining group exhibits oscillatory behavior around the fundamental apart from several price jumps). This suggests that for the small groups the news did not only have an influence on the individual forecasts, but also affected the market price dynamics.

4 Heuristic switching model

Empirical work on financial and housing price time series data and survey data on expectations shows that bubbles may be explained by non-rational expectations, such as trend-extrapolating forecasting rules (Coibion et al., 2018; Case et al., 2012; Barberis et al., 2018; Boswijk et al., 2007; Cornea-Madeira et al., 2017). In our market experiments, in stable markets subjects eventually coordinate expectations on the fundamental value, while in unstable markets the large bubbles seem to be amplified by coordination on trend-extrapolating expectations. To gain more insight in the impact of trend-extrapolating expectations in our experimental markets, in this section we fit the behavioral heuristics switching model of Anufriev and Hommes (2012) to our experimental data.

4.1 Model setup

Experimental laboratory data are often characterized by subjects heterogeneity, as e.g. stressed in the surveys of Arifovic and Duffy (2018) and Mauersberger and Nagel (2018). To model heterogeneity in expectations, Anufriev and Hommes (2012) developed a behavioural Heuristing Switching Model (HSM), an extension of Brock and Hommes (1997), and fitted the HSM to various learning to forecast experimental data sets[18].

The idea behind the HSM is that agents do not use a single forecasting rule, but they are heterogeneous in the rules they are using and switch between these rules, based on their relative performance. The average expected price in the market equals the weighted average of the expected prices produced by the heuristics: $\bar{p}^e_{t+1} = \sum_{i=1}^{4} n_{i,t} p^e_{i,t+1}$ 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. The HSM thus allows us to measure the impact of each of the forecasting rules. It is important to note that the rules used in HSM only use information that is available for subjects in the experiment. That is, expectations are formed based on observed realised prices and previous forecasts. To keep the model simple, Anufriev and Hommes (2012) use the following four rules:

Adaptive expectations: $p^e_{ADA,t+1} = 0.65\, p_{t-1} + 0.35\, p^e_{1,t},$

Weak trend-following rule: $p^e_{WTR,t+1} = p_{t-1} + 0.4\,(p_{t-1} - p_{t-2}),$

Strong trend-following rule: $p^e_{STR,t+1} = p_{t-1} + 1.3\,(p_{t-1} - p_{t-2}),$

Learning anchor & adjustment rule: $p^e_{LAA,t+1} = 0.5\,(p^{av}_{t-1} + p_{t-1}) + (p_{t-1} - p_{t-2}),$

[18]See also the recent work of (Anufriev et al., 2018) fitting a genetic algorithm model where agents select "smart" forecasting heuristics to various experimental data sets.

where p^{av}_{t-1} 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 (ADA) prices converge (slowly) monotonically to the fundamental price. Under weak trend-following rule (WTR) 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 (STR) the market is unstable and a large bubble occurs. Finally, under the learning anchor and adjustment (LAA) 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^{av}_{t-1} (a proxy for the long run equilibrium price). The LAA rule is the only rule able to predict turning points, consistent with bubble and crash oscillatory behavior[19].

As agents switch between these forecasting rules, we need to specify the switching process. In every period the decision rules are evaluated by a performance measure (U_i) that corresponds to how subjects are paid in the experiment: $U_{i,t-1} = -\big(p_{t-1} - p^e_{i,t-1}\big)^2 + \eta U_{i,t-2}$, 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 most recent forecasting error. Based on this performance measure, the evolution of fractions follows the discrete choice model with some inertia. In each period, a fraction δ of agents does not switch rules, whereas a fraction $1 - \delta$ follows an evolutionary 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 it 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.

4.2 Simulations of the experimental markets

In order to look at the impact of the different rules, we fit the HSM to the experimental data using one-period ahead forecast simulations. These simulations use exactly the same information (past prices and forecasts) as subjects had in the experiment.[20] To initialize the model, the first two prices in the

[19] Appendix B.2 estimates first-order forecasting heuristics (Eq. 9) of the same form as these 4 rules of the HSM. These estimation results confirm that many subjects use trend-following rules with an anchor that gives more weight to the last price observation.

[20] Note that we do not incorporate the possible news in the rules. However, if news has an effect, it may be captured by agents switching to another rule.

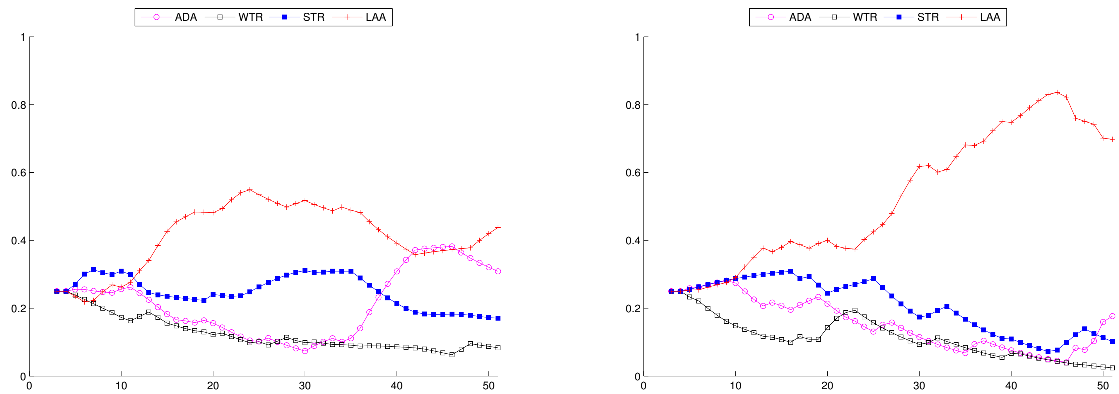


Figure 7: Average simulated fraction of rules for stable small (left) and large (right) markets

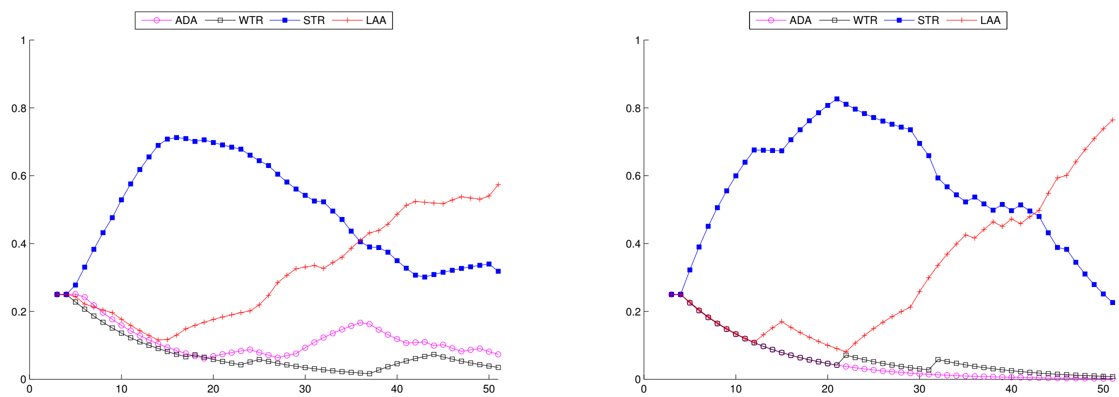


Figure 8: Average simulated fraction of rules for unstable small (left) and large (right) markets

experiment are used and for the initial forecast of the adaptive expectations rule we use $p^e_{ADA,3} = 50$ which was the midpoint of the interval we gave the subjects as a very likely price realisation for the first two periods. The initial shares of all forecasting rules are equal and fixed at $n_{i,3} = 0.25$ and $U_{i,3} = 0$. Following Anufriev and Hommes (2012), the parameters were fixed at $\beta = 0.4$, $\eta = 0.7$, and $\delta = 0.9$.[21]

Figures 7 and 8 show the average fractions of the different rules separately for the stable and unstable markets (with unstable markets defined again as those receiving news of overvaluation). In the stable markets there are some differences between small and large groups (Figure 7). What is common for both group-sizes is that the anchor and adjustment rule is dominant. In the large groups the LAA rule gradually

[21] Parameters are calibrated from earlier experimental results (Anufriev and Hommes, 2012). The results reported here are fairly robust w.r.t. the parameters β, η and δ. Furthermore, the results are fairly robust w.r.t. 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.

increases and dominates the market after 15 periods. The same is true for the small groups, but toward the end of the experiment adaptive expectations increases its share to $35-40\%$ about equalt to LAA. Apparantly, the large stable groups exhibit a little more fluctuations than the small stable groups. Note however, that in the more stable markets the difference in the performance of the rules is rather small, so it is hard to draw strong conclusions here.

For the unstable markets the simulated fractions are very similar for both small and large groups (Figure 8). In the unstable markets the strong trend-following rule dominates the market in the first 15-20 periods with a maximum share of $70-80\%$. Hence, according to the HSM the first 15-20 periods of the unstable markets, characterized by large bubbles, coincide with coordination on the strong trend-following rule. According to the HSM, bubbles are therefore amplified by strong trend-extrapolation. After the initial 15-20 periods, due to market crashes, the anchor and adjustment rule gains impact, and the strong trend-following rule's weight decreases after period 20. The coordination on the strong trend-following rule thus amplifies bubbles in these markets. High prices are expected, and these expectations are 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 summary, the behavioral HSM supports the hypothesis that bubbles are amplified by coordination of expectations on a strong trend-extrapolating rule.

5 Conclusion

A common objection to financial market and macroeconomic experiments is that laboratory markets are (too) small and therefore individual decisions may have a stronger influence on aggregate outcomes than in real markets. In this paper we study expectation formation and coordination on asset bubbles in large (about 100 participants) and small (6 participants) experimental markets. We add further realism to the markets by allowing for information contagion, that is, providing news from experts about market fundamentals. When the price moves far away from its fundamental value, we randomly choose some of the market participants to receive news about the market being either over- or undervalued. An important question then is whether in a large group these news incidents will spread and be successful in breaking the coordination on a bubble and drive prices back to the fundamental.

Our experiment reveals that coordination on bubbles is not sensitive to group size. In fact, our experiment reveals no substantial difference between small and large groups in the asset pricing experimental framework. For both small and large groups three different types of aggregate behaivour emerge: relatively stable markets, markets with moderately large bubbles and markets with very large bubbles. The news did not stabilise prices in all markets, as some bubbles only crashed after reaching an artificial upper bound imposed on prices. However, subjects reacted significantly to news: those exposed to the news of overvaluation predict lower prices compared to those who have not seen the news. These individual effects may lead to information contagion. For some, but not all, markets, both for small and large groups, news seem to break the coordination on bubbles and trigger market crashes.

An advantage of large groups is that it provides a more accurate description of bubble formation and crashes. Initially expectations are heterogeneous, but subjects quickly coordinate expectations. This may lead to above fundamental prices and subjects coordinating expectations on an initially weak, but later on stronger trend. As the bubble continues to grow, heterogeneity in expectations gradually increases and the bubble eventually crashes, either due to the contagion of the news or by reaching an exogenous upper bound. After a market crash heterogeneity peaks to very high levels and continues to stay high representing a state of confusion. This pattern of bubble formation and market crash is clear and smooth in large markets. In small markets the pattern is similar, but it appears noisier.

To further investigate individual behaviour and coordination on trend-extrapolation, we fitted a behavioural heuristic switching model to the observed experimental data. This model uses four simple forecasting rules that subjects might use in the experiment to form their expectations. These rules are adaptive expectations, a weak and a strong trend-following rule, and an anchoring and adjustment rule. Fitting the heuristic switching model to the experimental data gives insight in the relative impact of each of these rules. For markets that exhibit large bubbles, the strong trend-followers dominate the market, especially in the first 20-30 periods, with a peak around $70 - 80\%$. This rule performs the best in an environment where prices increase rapidly, which can lead to a majority using this rule. Coordination on trend-following expectations thus reinforces further price increases and bubbles become unavoidable.

To sum up, in learning-to-forecast asset pricing experiments the results for small groups, such as coordination on bubbles, carry over to large groups. This lends support to the many learning-to-forecast experiments with small groups in the literature. However, this robustness result cannot be extrapolated to all macroeconomic or financial market experiments. In other environments aggregate behaviour in large groups may or may not differ from small group behaviour. Designing large group experiments in macro and finance is an important area of research to test the robustness of small group behaviour and/or the emergence of new phenomena arising in large macro-financial systems.

References

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.

Anufriev, M., Hommes, C. H., and Makarewicz, T. (2018). Simple forecasting heuristics that make us smart: evidence from different market experiments. *Journal of the European Economic Association*, forthcoming.

Arifovic, J. and Duffy, J. (2018). Heterogeneous agent modeling: Experimental evidence. In Hommes, C. and LeBaron, B., editors, *Handbook of Computational Economics, Volume 4: Heterogeneous agent modelling*, pages 491–540. Elsevier, Amsterdam.

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

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.

Boswijk, H. P., Hommes, C. H., and Manzan, S. (2007). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, 31(6):1938–1970.

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*, 65(5):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.

Case, K. E., Shiller, R. J., and Thompson, A. K. (2012). What have they been thinking? homebuyer behavior in hot and cold markets. *Brookings Papers on Economic Activity*, 2012(Fall):265–298.

Coibion, O., Gorodnichenko, Y., and Kamdar, R. (2018). The formation of expectations, inflation, and the phillips curve. *Journal of Economic Literature*, 56:1447–1491.

Cornea-Madeira, A., Hommes, C., and Massaro, D. (2017). Behavioral heterogeneity in us inflation dynamics. *Journal of Business & Economic Statistics*, pages 1–13.

Duffy, J. (2016). Macroeconomics: A Survey of Laboratory Research. In *Handbook of Experimental Economics, Volume 2*, pages 1–90. Princeton University Press.

Evans, G. W. and Honkapohja, S. (2001). *Learning and expectations in macroeconomics.* Princeton University Press.

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. (2019). The strategic environment effect in beauty contest games. *Games and Economic Behavior*, in press.

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. (2018). Monetary and fiscal policy design at the zero lower bound–evidence from the lab. *Economic Inquiry*, accepted.

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. (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*.

Lucas, R. E. (1972). Econometric testing of the natural rate hypothesis. In Eckstein, O., editor, *The Econometrics of Price Determination [Conference]*, pages 50–59, Washington DC. Board of Governors of the Federal Reserve System and Social Science Research Council.

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(1):131–162.

Mauersberger, F. and Nagel, R. (2018). Heterogeneity in (micro) experiments or experimental games. In Hommes, C. and LeBaron, B., editors, *Handbook of Computational Economics, Volume 4: Heterogeneous agents models*, pages 541–634. Elsevier, Amsterdam.

Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica*, 29(3):315–335.

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

Noussair, C. N. and Tucker, S. (2013). Experimental research on asset pricing. *Journal of Economic Surveys*, 27(3):554–569.

Sargent, T. J. (1993). *Bounded rationality in macroeconomics*. Oxford University Press.

Smith, V. L., Suchanek, G. L., and Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica*, 56(5):1119–1151.

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

Figure 9: News-item when the stock is overvalued

In this appendix an example of the news (see Figure 9) 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.

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.

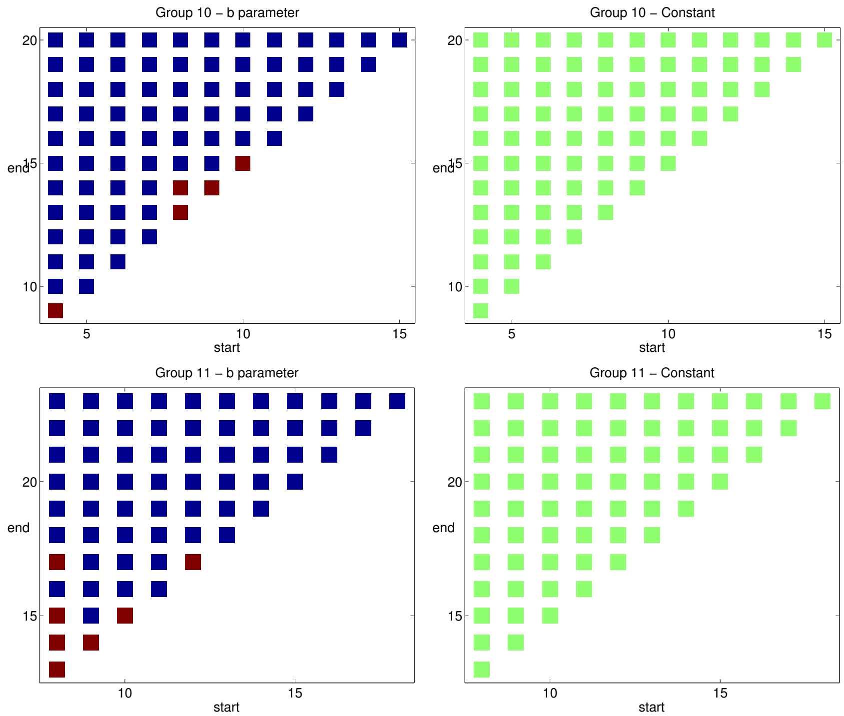
Payoff Table

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 10: Regression result of estimating anchoring on the price (Eq. (7)) on small groups

Appendix B Supplementary analysis, figures and tables

B.1 Bubble-growth

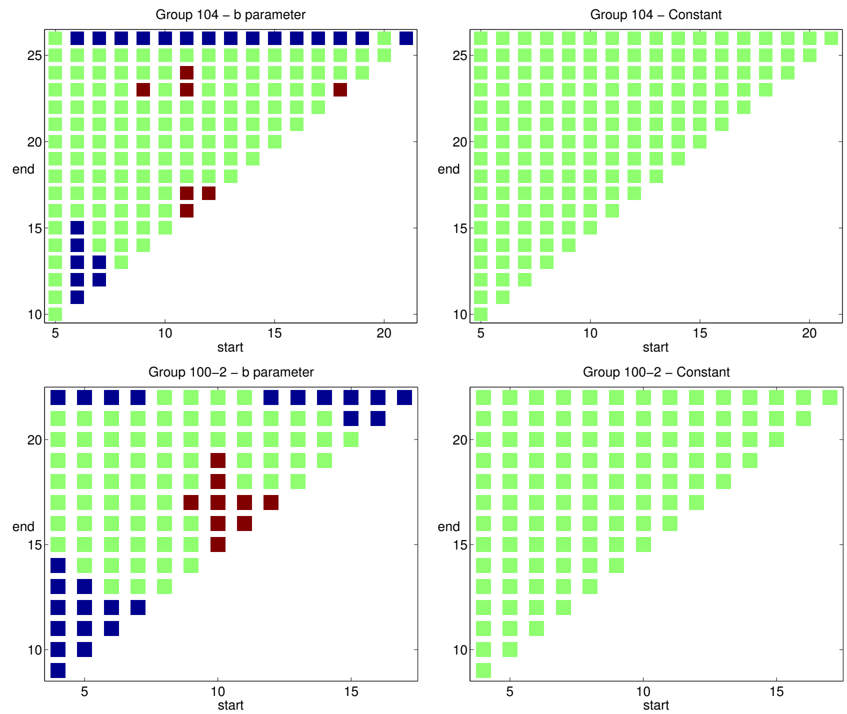
Figure 1 suggests that for the large bubbles the price increases with a higher than exponential growth rate. Following Hüsler et al. (2013) we estimate the growth rate for both the small and the large groups with two specifications. The first specification assumes achoring on the price, and uses the following equation:

$$\log\left(\frac{\bar{p}_t}{\bar{p}_{t-1}}\right) = a_1 + b_1\bar{p}_{t-1} \tag{7}$$

The second specification is based on anchoring on return:

$$\log\left(\frac{\bar{p}_{t+1}}{\bar{p}_t}\right) = a_2 + b_2 \log\left(\frac{\bar{p}_t}{\bar{p}_{t-1}}\right). \tag{8}$$

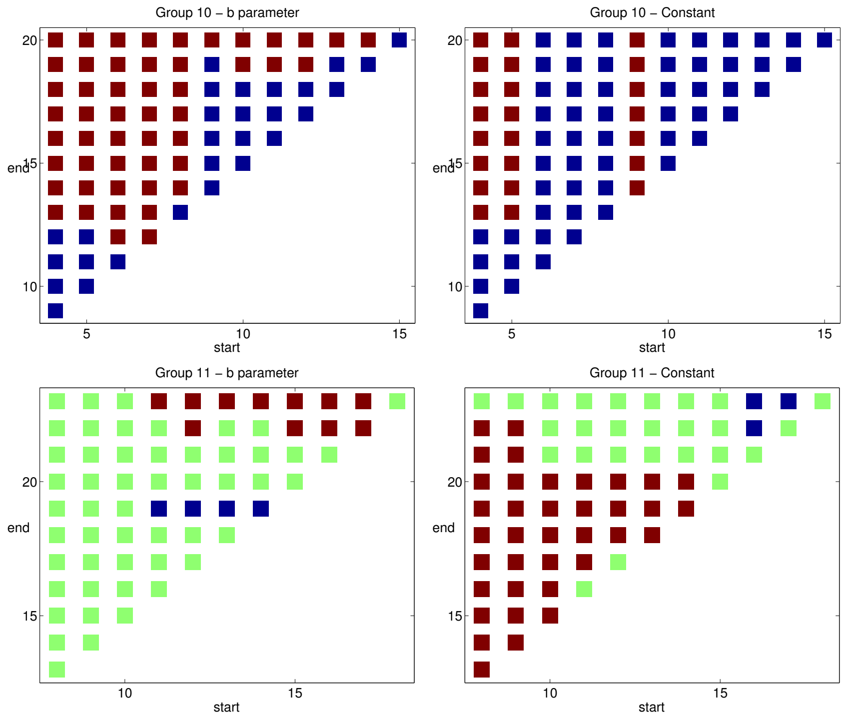
In both cases, if $a_i > 0$ and $b_i > 0$ ($b_i < 0$), then the growth is larger (smaller) than exponential, but the feedback is based on prices in the first, and on return in the second specification. Looking at Figure 1 again, we can see two large and two small markets with very large bubbles. These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



Notes: See explanation under Figure 10.

Figure 11: Regression result of estimating anchoring on the price (Eq. (7)) on large groups

different for the 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 10-13 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 starting and ending periods, whereas the growth rate is faster in the large group (but still not significantly faster than exponential growth in most periods). This can also be seen in Figure 1. However, if we look at Figures 12 and 13 there is no clear difference between group sizes. Anchoring on return results in a faster than exponential growth in both markets for early starting rounds, but in a slower than exponential growth rate towards the peak. Also, in this specification the parameter a_2 is significantly positive for early starting points. These results suggest that the growth rate might be higher in the large groups than in the small groups, but we cannot draw strong conclusions given the low number of observations.



Notes: See explanation under Figure 10.

Figure 12: Regression result of estimating anchoring on the returns (Eq. (8)) on small groups

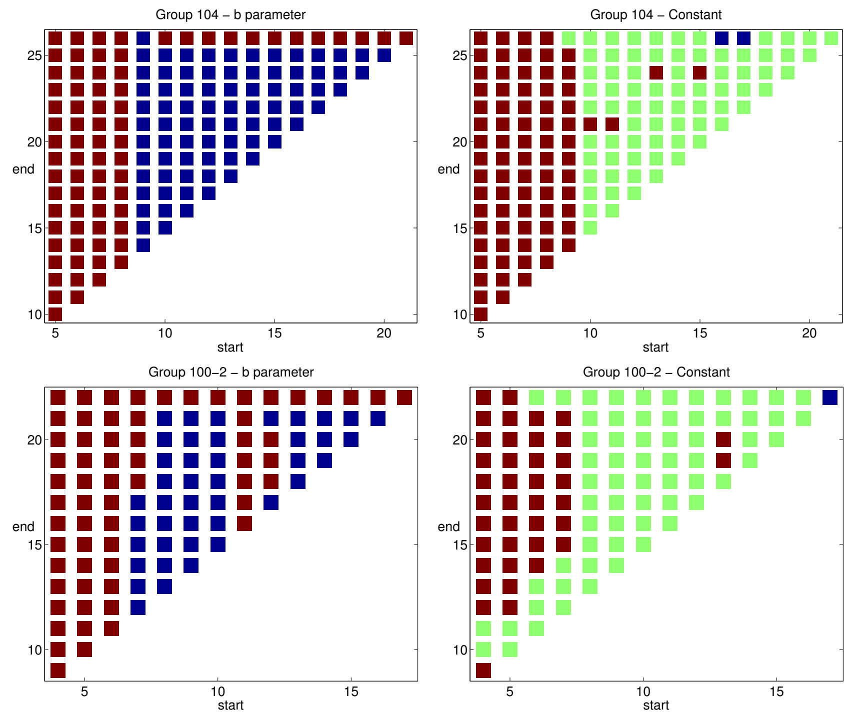
B.2 Individual behaviour

Figures 14 and 15 show the individual forecasts and the market price in each market. Figures 16 and 17 show the market price and the standard deviation of individual forecasts in each market.

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 a simple first-order anchor and adjustment heuristics (see e.g. Anufriev et al., 2018) which is given by the following equation:

$$p^e_{i,t+1} = \alpha p_{t-1} + (1-\alpha)p^e_{i,t} + \beta(p_{t-1} - p_{t-2}) + v_t, \tag{9}$$

This first-order heuristics forecasting rule is relatively simple but can capture different heuristics. For $\beta = 0$ the rule reduces to adaptive expectations. Taking $\alpha = 1$, and $\beta = 0$ we get the naive expectations rule. More generally, the rule is in the form of an anchoring and adjusting rule (see also Section 4) with an anchor using a weighted average of the last observed market price and the last own forecast. From



Notes: See explanation under Figure 10.

Figure 13: Regression result of estimating anchoring on the returns (Eq. (8)) on large groups

this anchor the forecast 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 (9) is estimated, after removing outliers and filling up missing data (using linear interpolation).[22] The first 5 periods are disregarded, to allow for a short learning phase. We estimate the model twice for each individual. First, 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 175 of the 676 (26%) subjects survive all three model-specification tests at the 5%-level.

Table 3 summarises the average coefficients of the regressions. Panel A presents the average coefficients over all 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

[22] An 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*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).

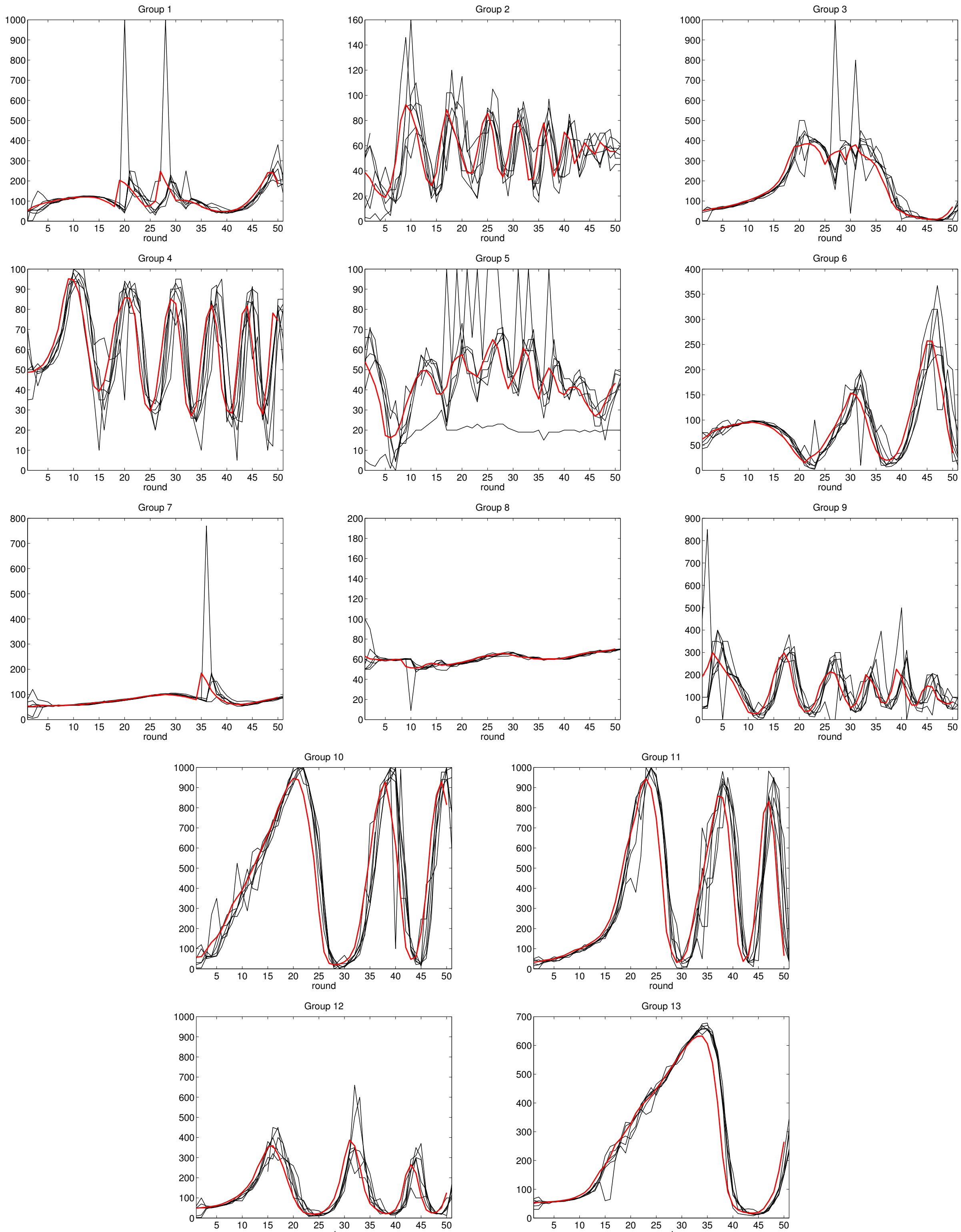


Figure 14: Individual predictions in small markets

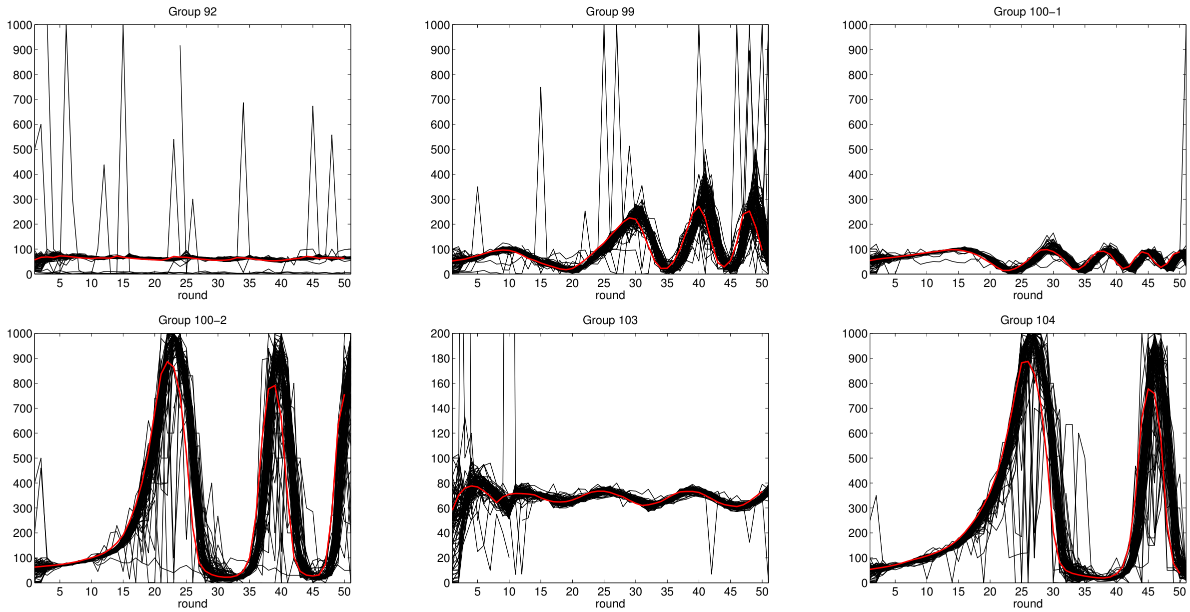


Figure 15: Individual predictions in large markets

coefficients 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, with a higher weight on the latter. Furthermore, on average they seem to be trend extrapolators, with $0 < \beta < 1$. The estimated coefficients are similar for the different group sizes. However, comparing stable and unstable groups, we see that subjects are stronger trend-followers in the unstable markets, that is, they have a higher estimated β coefficient. This difference is only significant for small groups ($p = 0.02$ with a ranksum test on market level) in Panel A. Unfortunately we have too few observations of large markets (3 stable and 3 unstable) to find significant differences. All the other differences are insignificant at the 10%-level.

In the small groups 22 out of the 78 subjects' behaviour can be described by first-order heuristics. These 22 subjects are distributed among 11 groups. These groups consist of both stable and unstable markets. 12 out of the 22 participants have $\alpha > 0.75$ which suggest an anchor that is mainly based on the last observed price (corresponding to naive expectations if $\beta = 0$). All but one subject has a significantly positive β parameter, ranging from 0.39 (Gr. 11) to 1.74 (Gr. 6). Both these extrema are observed in unstable markets. 4 subjects have a pure trend-following rule with $\alpha \in (0.9, 1.1)$ 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 27 out of the 153 subjects have $\alpha \in (0.9, 1.1)$ with 3 subjects having $\beta = 0$ as well. 18 subjects have an insignificant β coefficient, all other individuals are trend-following with β ranging from 0.30 (Gr. 103 - stable) to 1.82

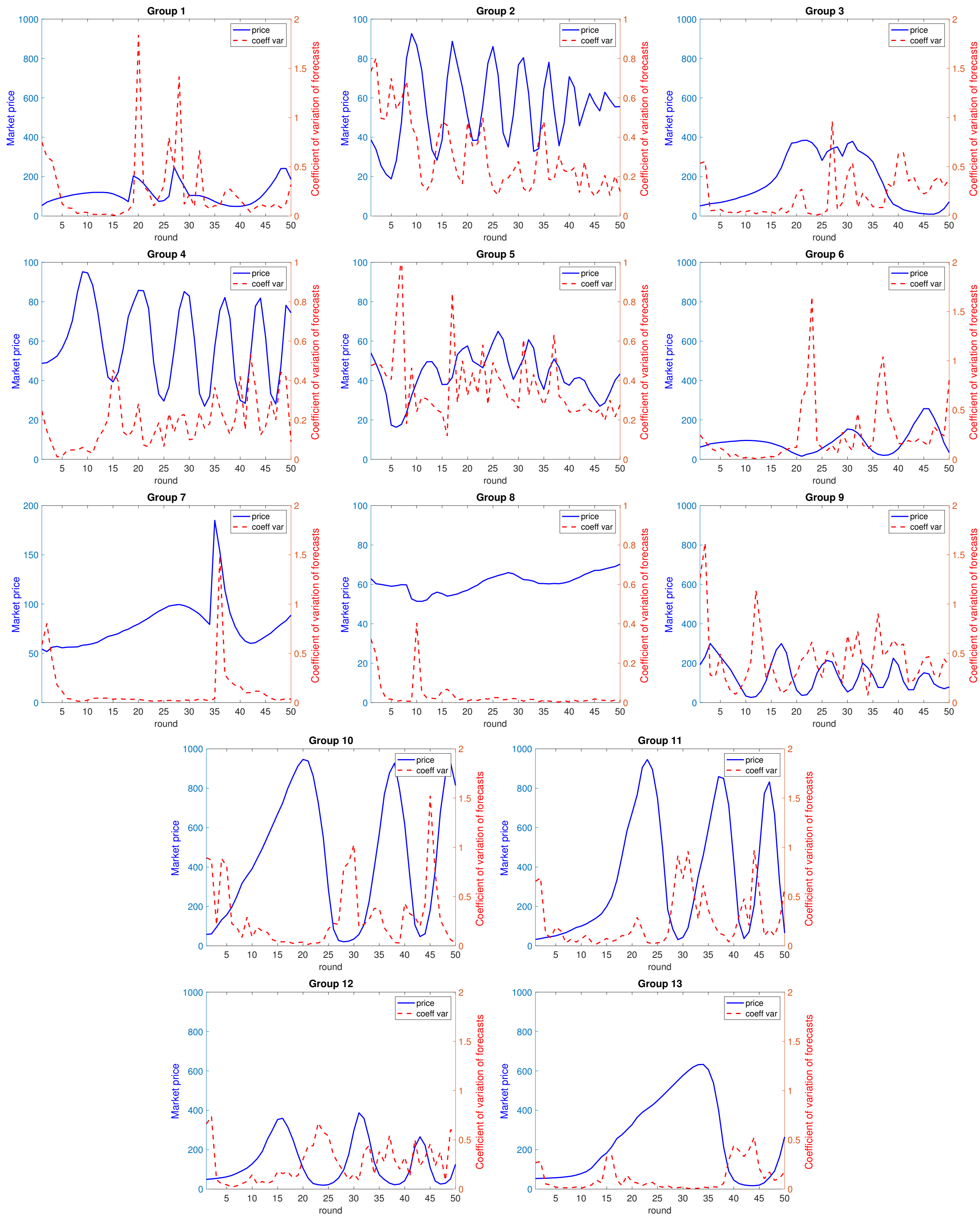


Figure 16: Realized prices and coefficient of variation of individual forecasts in the 13 small markets.

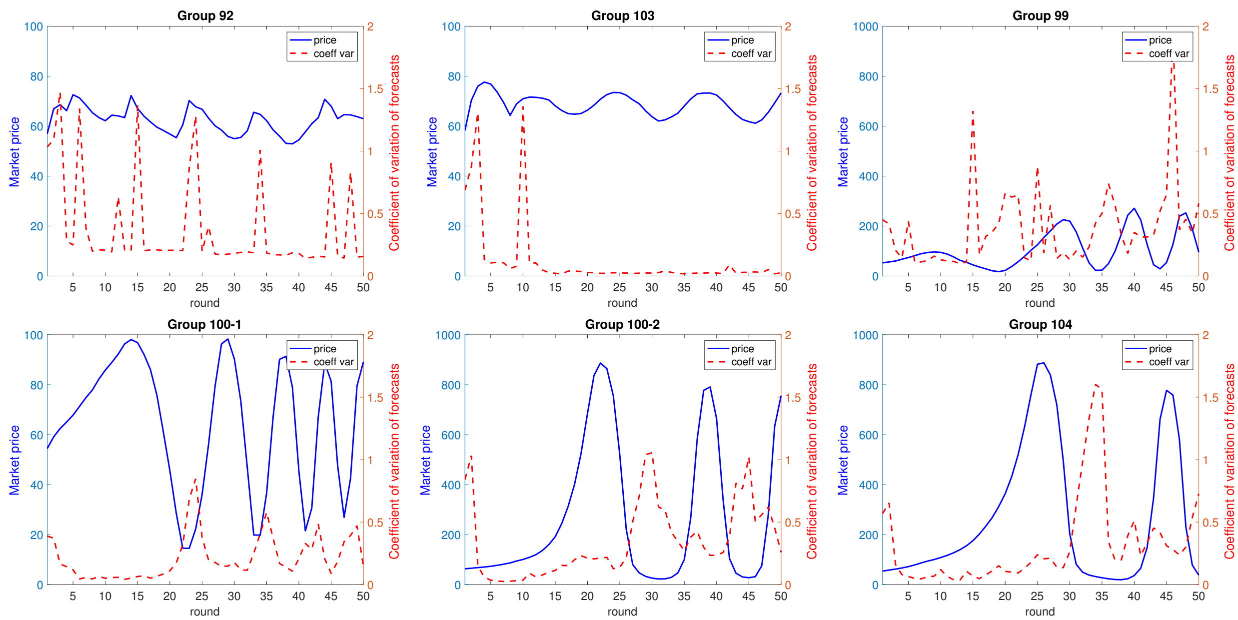


Figure 17: Realized prices and coefficient of variation of individual forecasts in the six large markets.

(Gr. 99 - unstable).

To summarise this section, about 26% of our subjects' behaviour can be described by the first-order heuristics given by Eq. (9). 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 a weighted average of the last observed price and the last own forecast, with more weight to the last observed price. 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.[23] These switches cannot be described by this simple rule and are the topic of the next section.

B.3 Heuristic Switching Model

Figures 18-20 show the HSM simulated for the three different behavioural patter 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

[23]In 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).

Market type	fraction of subject	$\bar{\alpha}$	β
Panel A: All models			
Small pooled		0.68 (0.39)	0.70 (0.38)
Small stable		0.62 (0.41)	0.49 (0.35)
Small bubble		0.72 (0.38)	0.83 (0.34)
Large pooled		0.77 (0.52)	0.76 (0.38)
Large stable		0.71 (0.49)	0.70 (0.43)
Large bubble		0.83 (0.54)	0.82 (0.33)
Panel B: Good models - significant coefficients			
Small pooled	28% (22/78)	0.91 (19)	0.81 (21)
Small stable	17% (5/30)	0.96 (5)	0.54 (4)
Small bubble	35% (17/48)	0.89 (14)	0.86 (17)
Large pooled	26% (153/598)	0.88 (141)	0.81 (135)
Large stable	33% (98/295)	0.79 (91)	0.77 (83)
Large bubble	18% (55/303)	1.3 (50)	0.87 (52)

Notes: Markets are divided into stable and bubble markets in the same way as in Table **??**. 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 3: Summary of first-order heuristics

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 patterns qualitatively. To quantify the model performance, we look at different benchmark models with and without heterogeneity, and determine the mean-squared error (MSE) of these models by calculating the average squared difference between the simulated and the experimental price. We consider 6 different homogeneous rules: the four rules we have used in the HSM, plus the fundamental rule ($p^e_{t+1} = p^f$ for all t) and the naïve expectations rule ($p^e_{t+1} = p_{t-1}$). 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.

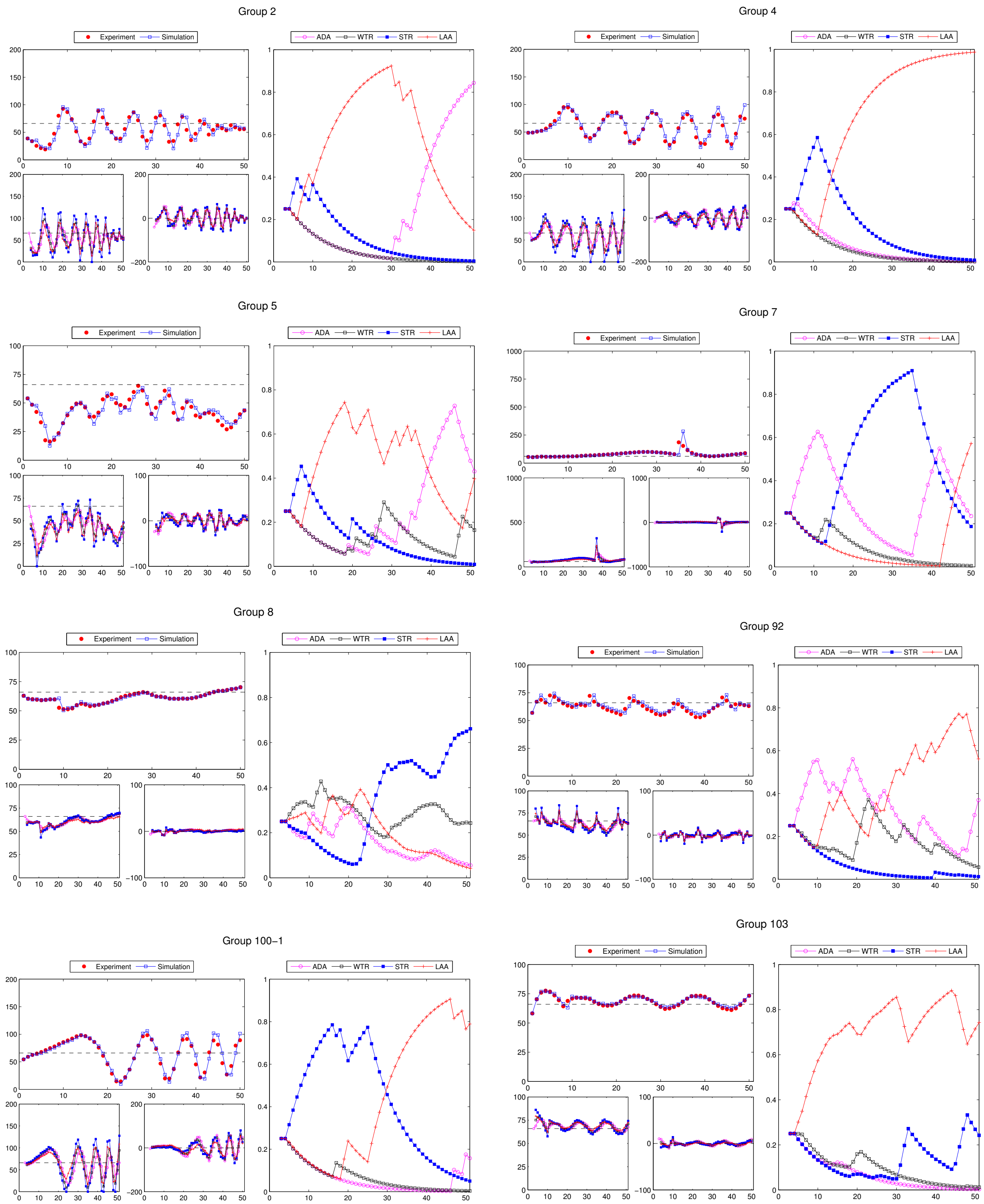


Figure 18: HSM for stable markets in small (first 5 panels) and large (last 3 panels) groups.

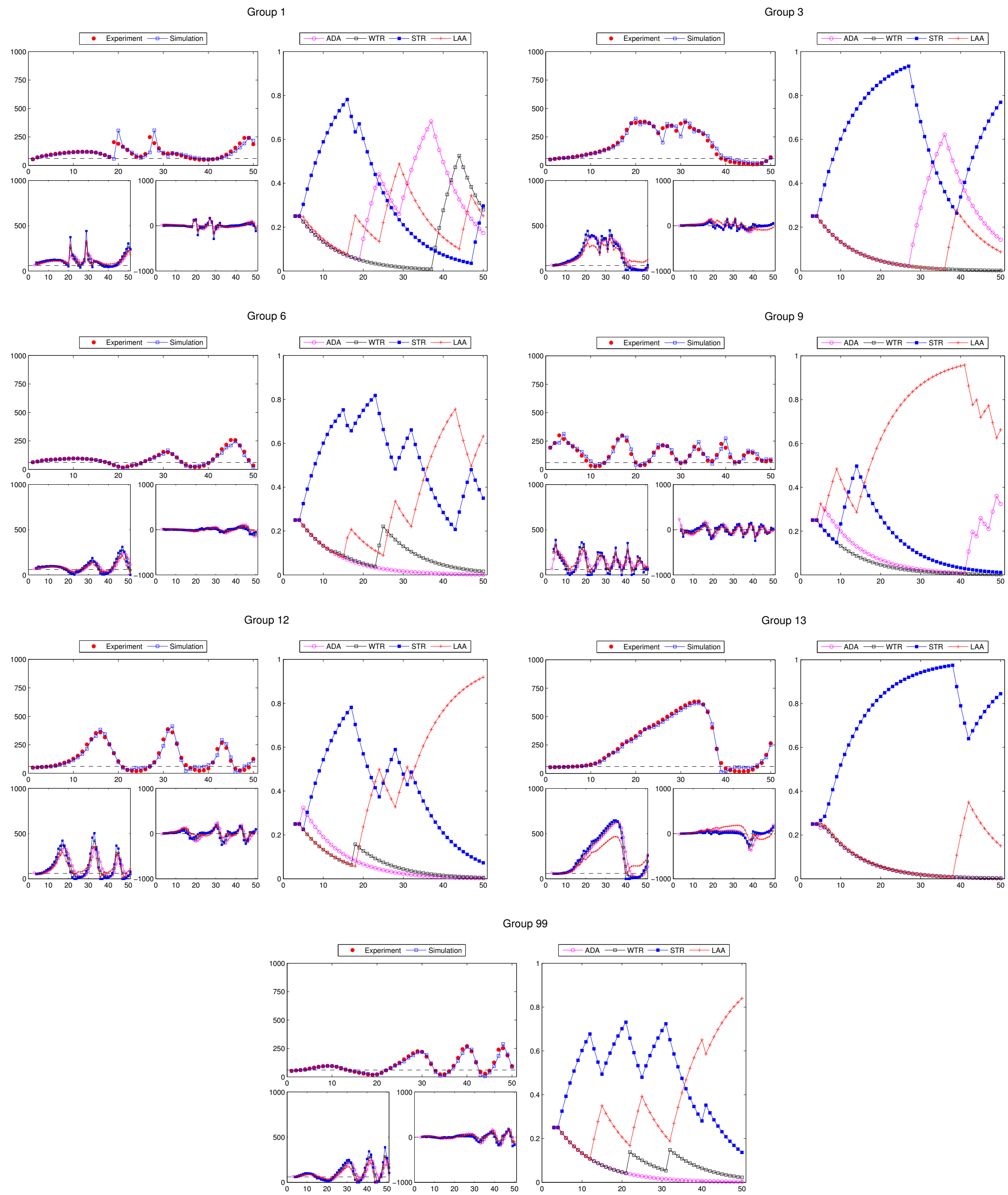


Figure 19: HSM for large bubbles in small (first 6 panels) and large (last panel) groups.

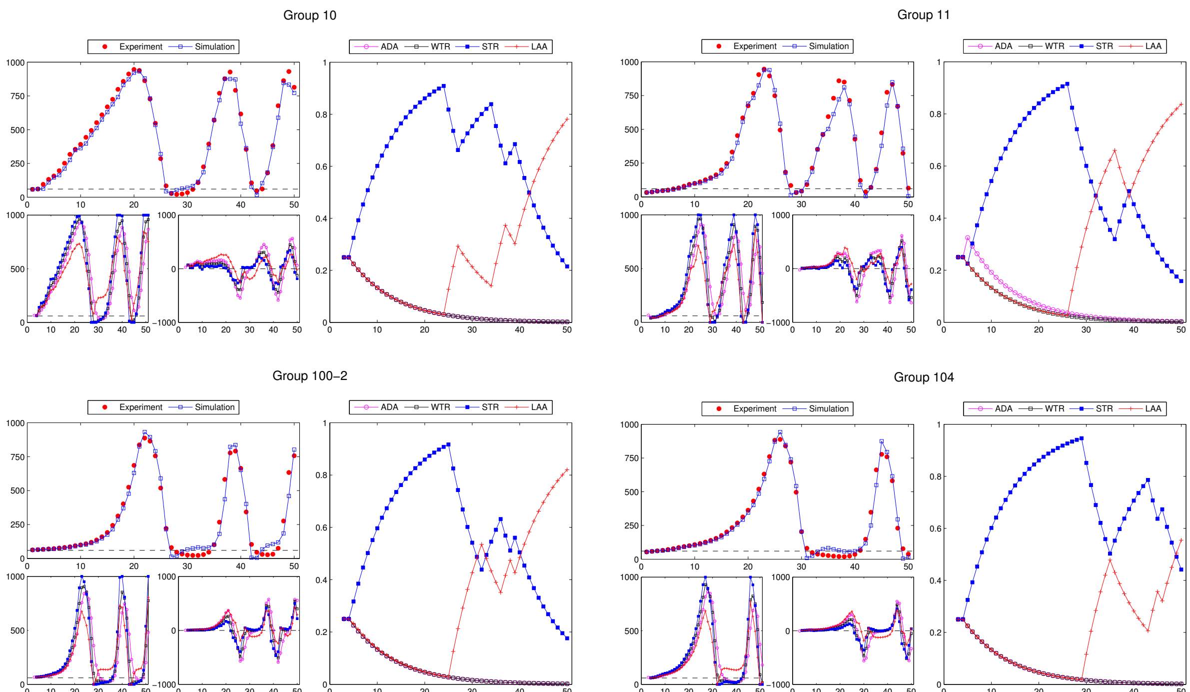


Figure 20: HSM for very large bubbles in small (upper) and large (lower panels) groups.

	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 4.2 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 4: Mean squared error for the different heuristics

Table 4 lists the MSE for each market for each model, and the parameters for the best-fitting HSM-model.[24] 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 are 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 unstable markets with bubbles. There the variance in performance of the different rules are much larger as well. In none of the models the fundamental rule is the best. The fundamental rule produces a high MSE, even in the stable markets. If we turn to the heteregoneous rules,

[24] In 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.

we observe that in most cases (17 out of 19 markets) the benchmark HSM performs the best (after the fitted HSM which is the best because it nests all other models). There are two markets in which the fixed fraction performs better. In these cases the benchmark HSM has parameters that are quite different from the optimal ones, which can cause a decline in the performance of the benchmark HSM. In most of the markets, allowing for heterogeneity, and moreover allowing for switching substantially improves the fit. Subjects do not seem to stick to one forecasting rule if it proves to be inefficient.

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 agents behave as if they were learning and switching to the optimal rule very quickly, in some other markets they learn at 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 of Anufriev and Hommes (2012). This suggest that our subjects give more weight to more recent observations and are willing to switch more frequently than in the benchmark model.