Online appendix

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

Cars Hommes[1,2,3], Anita Kopányi-Peuker[4,1] and Joep Sonnemans[1,2]

[1]University of Amsterdam

[2]Tinbergen Institute

[3]Bank of Canada

[4]CPB Netherlands Bureau for Economic Policy Analysis

June 23, 2020

Appendix A Experimental instructions

Figure 1: News-item when the stock is overvalued

In this appendix an example of the news (see Figure 1) 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		

Appendix B Supplementary analysis, figures and tables

B.1 Bubble-growth

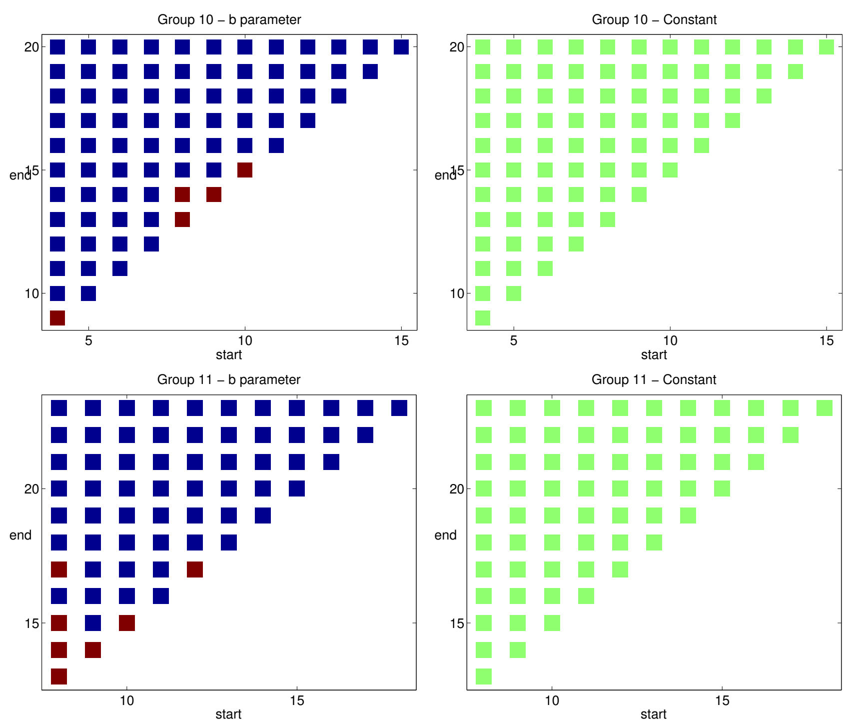
Figure 1 suggests that for the large bubbles the price increases with a higher than exponential growth rate. Following Hüsler et al. (2013) we estimate the growth rate for both the small and the large groups with two specifications. The first specification assumes achoring on the price, and uses the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

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

The second specification is based on anchoring on return:

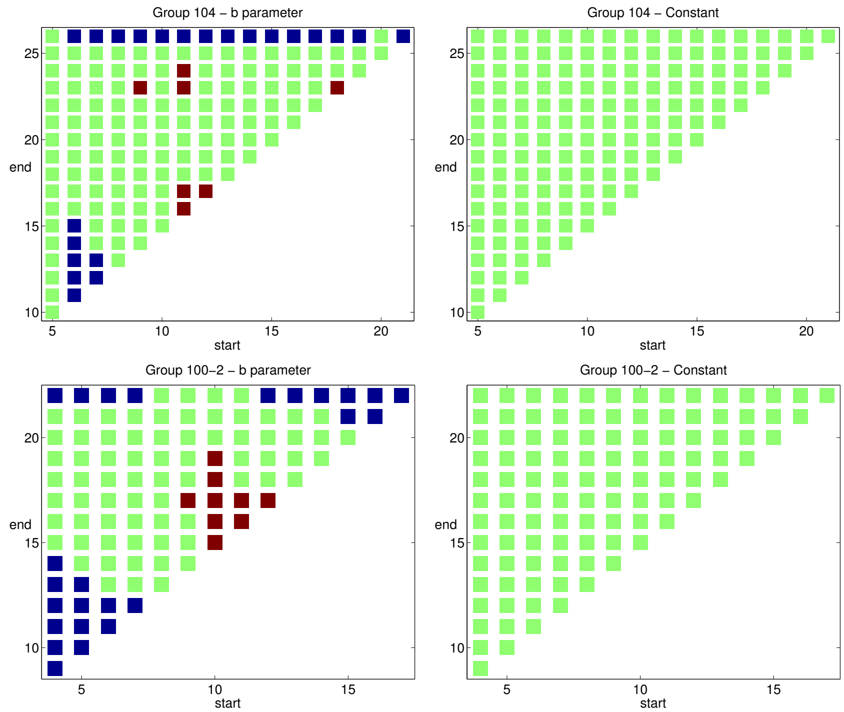
$$\log\left(\frac{\bar{p}_{t+1}}{\bar{p}_t}\right) = a_2 + b_2 \log\left(\frac{\bar{p}_t}{\bar{p}_{t-1}}\right). \tag{2}$$

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



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.

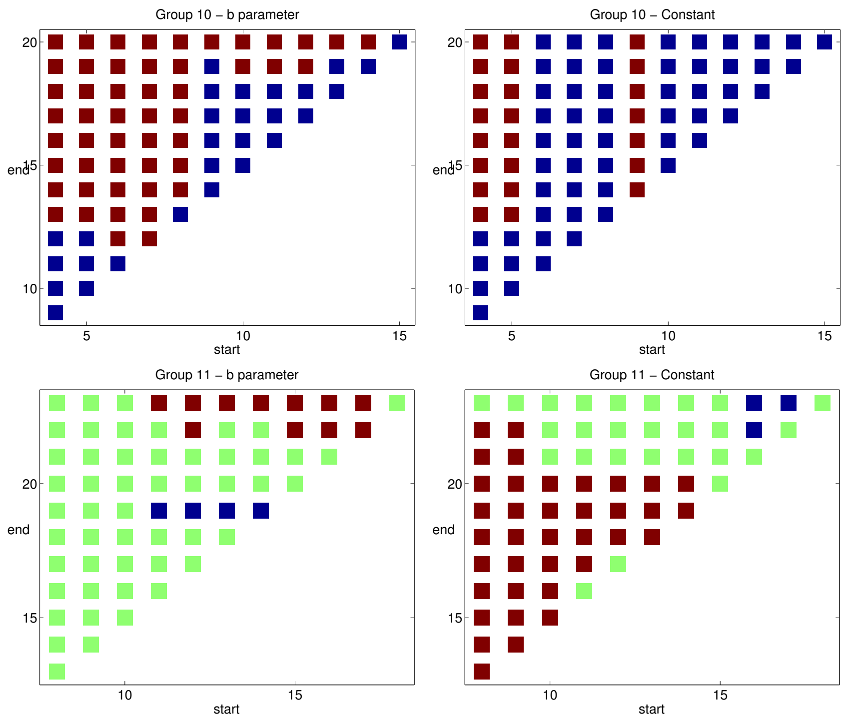
Figure B.1: Regression result of estimating anchoring on the price (Eq. (1)) on small groups



Notes: See explanation under Figure B.1.

Figure B.2: Regression result of estimating anchoring on the price (Eq. (1)) on large groups

11 for the small groups, and markets 100-2 and 104 for the large groups. The bubble periods are different for the different groups. The starting period is when the price exceeds the fundamental price 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 B.1-B.4 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 B.3 and B.4 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 B.1.

Figure B.3: Regression result of estimating anchoring on the returns (Eq. (2)) on small groups

B.2 Market prices and predictions in all markets

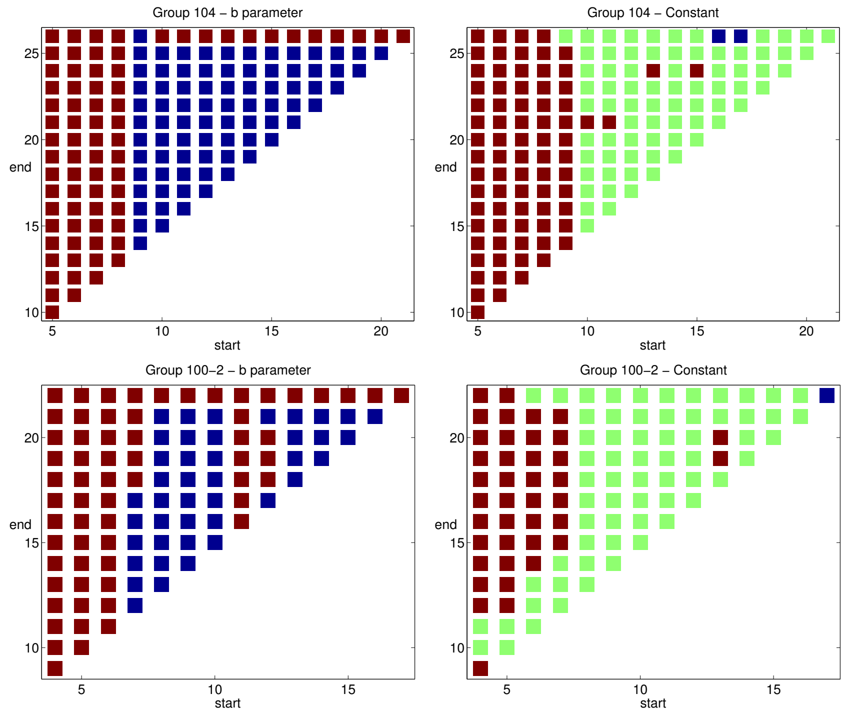
Figures B.5 and B.6 show the individual forecasts and the market price in each market. Figures B.7 and B.8 show the market price and the standard deviation of individual forecasts in each market.

B.3 Individual behaviour

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 numbers 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., 2019) which is given by the following equation:

$$p^e_{i,t+1} = \alpha p_{t-1} + (1-\alpha)p^e_{i,t} + \beta(p_{t-1} - p_{t-2}) + v_t, \tag{3}$$

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



Notes: See explanation under Figure B.1.

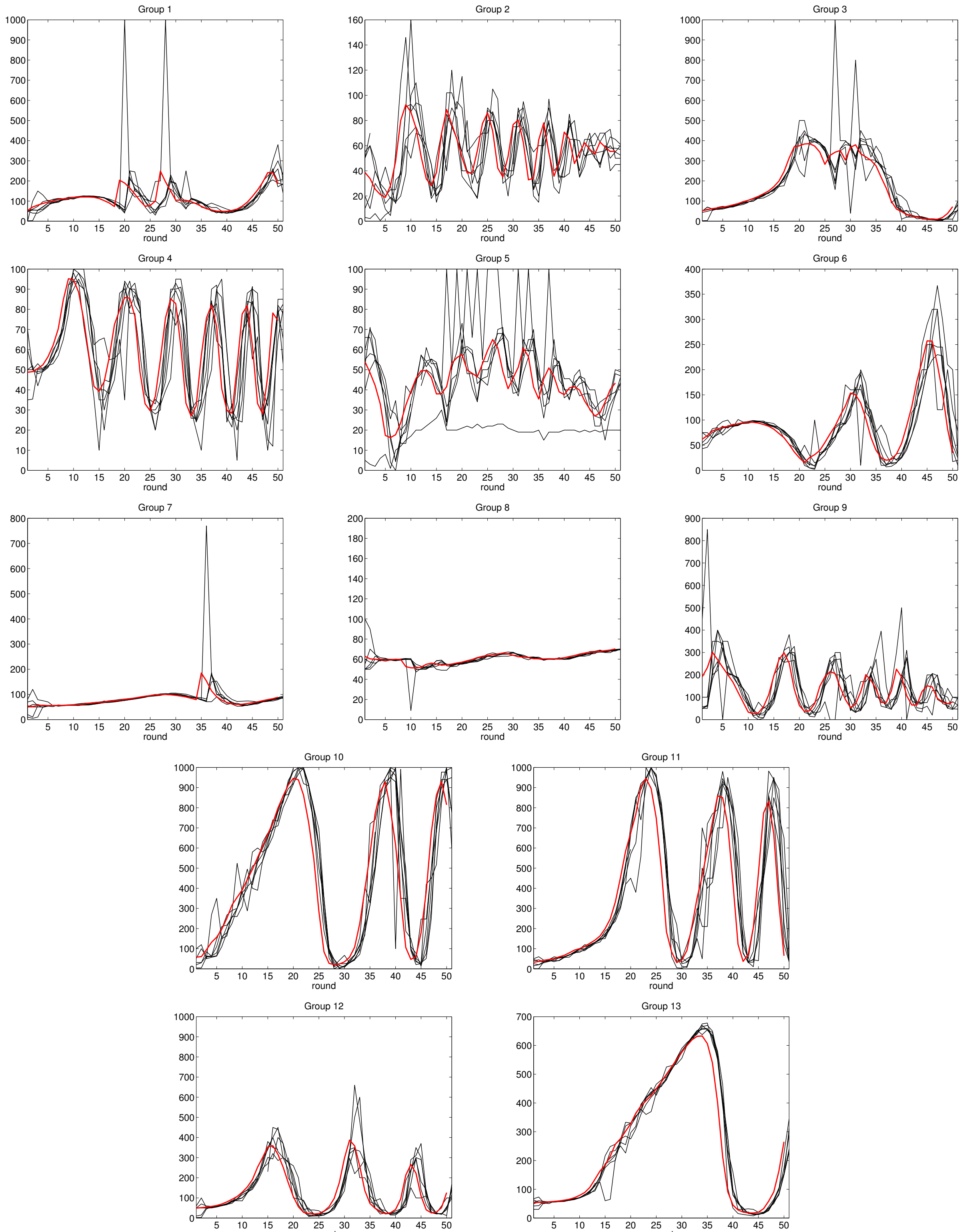
Figure B.4: Regression result of estimating anchoring on the returns (Eq. (2)) on large groups

rule. More generally, the rule is in the form of an anchoring and adjusting rule (see also Section B.4) with an anchor using a weighted average of the last observed market price and the last own forecast. From 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 (3) is estimated, after removing outliers and filling up missing data (using linear interpolation).[1] 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.

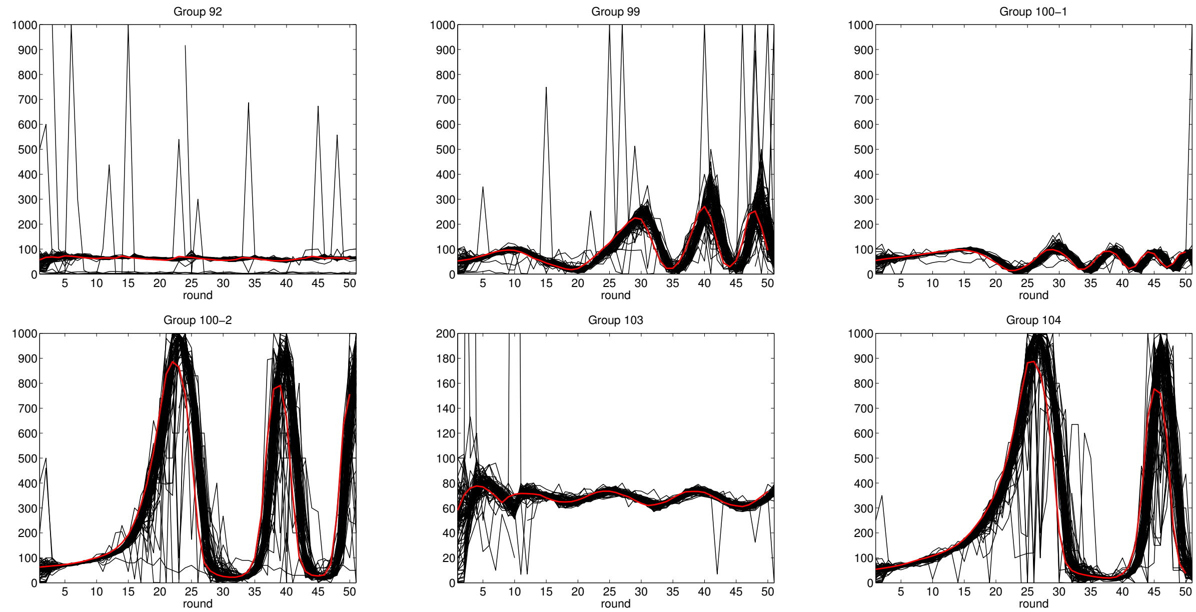
Table B.1 summarises the average coefficients of the regressions. Panel A presents the average coefficients

[1] 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).



Notes: The price is depicted on the $y-$axis. Each plot corresponds to a different market. Each black line denotes an individual prediction, whereas the red line corresponds to the market price.

Figure B.5: Individual predictions in small markets



Notes: The price is depicted on the $y-$axis. Each plot corresponds to a different market. Each black line denotes an individual prediction, whereas the red line corresponds to the market price.

Figure B.6: Individual predictions in large markets

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 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

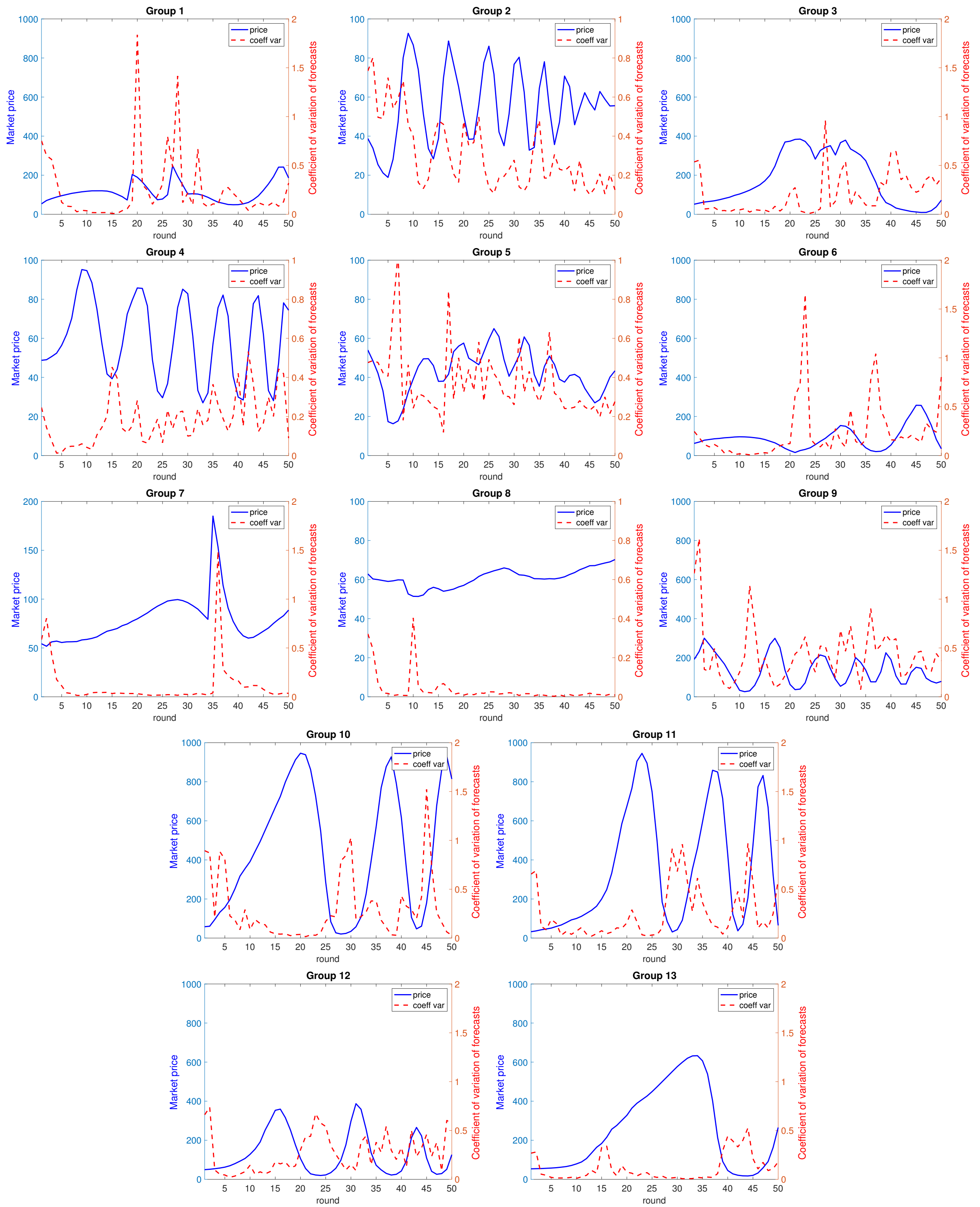


Figure B.7: Realized prices and coefficient of variation of individual forecasts in the 13 small markets.

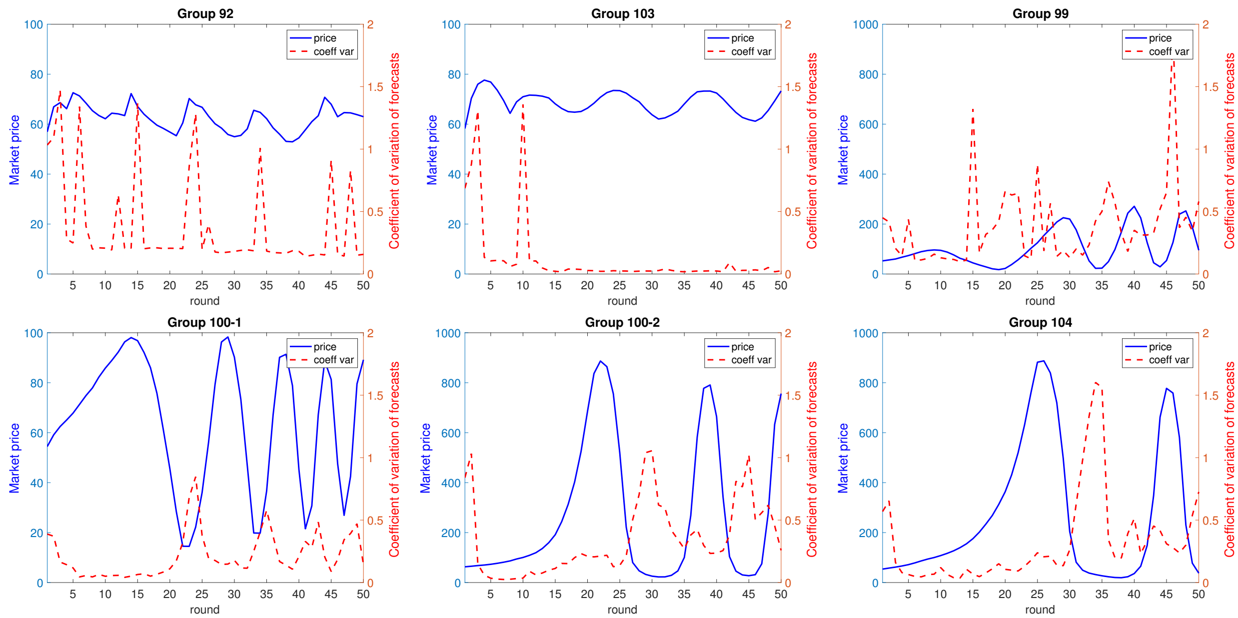


Figure B.8: Realized prices and coefficient of variation of individual forecasts in the six large markets.

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 (Gr. 99 - unstable).

To summarise this section, about 26% of our subjects' behaviour can be described by the first-order heuristics given by Eq. (3). 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.[2] These switches cannot be described by this simple rule and are the topic of the next section.

B.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.,

[2]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 described in Sect. 3.1. 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 B.1: Summary of first-order heuristics

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.

B.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 Heuristic Switching Model (HSM), an extension of Brock and Hommes (1997), and fitted the HSM to various learning to forecast experimental data sets[3].

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

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 price. 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[4].

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: $U_{i,t-1} = -\left(p_{t-1} - p^e_{i,t-1}\right)^2 + \eta U_{i,t-2}$, 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

[3]See also the recent work of (Anufriev et al., 2019) fitting a genetic algorithm model where agents select "smart" forecasting heuristics to various experimental data sets.

[4]Appendix B.3 estimates first-order forecasting heuristics (Eq. 3) 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.

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.

B.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.[5] To initialize the model, the first two prices in the 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 $\delta = 0.9$.[6]

Figures B.9 and B.10 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 B.9). What is common for both group-sizes is that the anchor and adjustment rule is dominant. In the large groups the LAA rule gradually 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 equal to LAA. Apparently, 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.

[5]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.

[6]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.

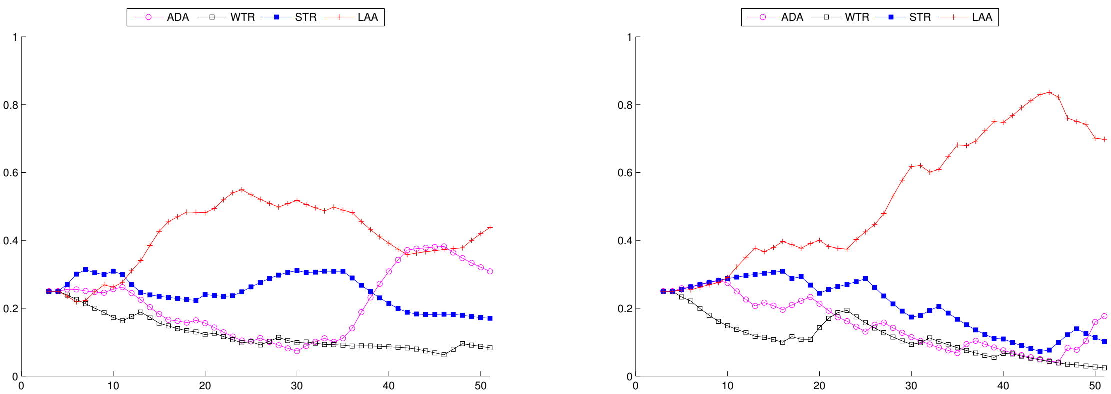


Figure B.9: Average simulated fraction of rules for stable small (left) and large (right) markets

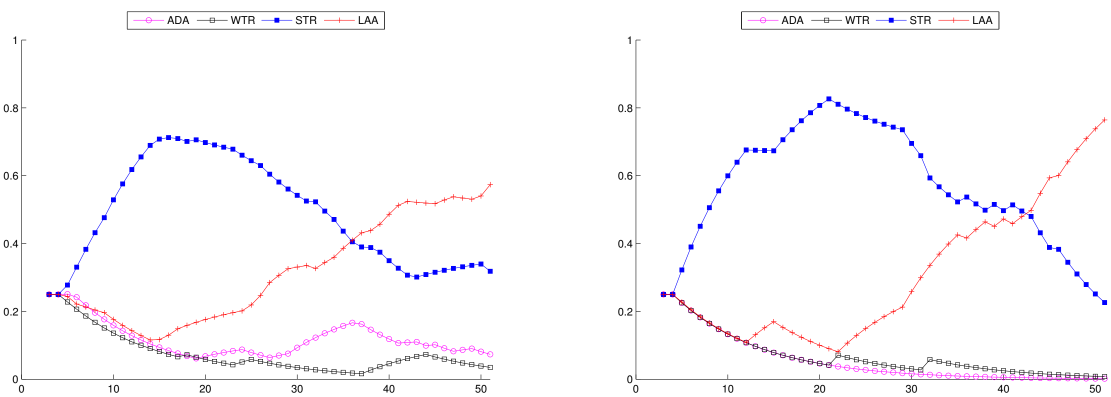


Figure B.10: Average simulated fraction of rules for unstable small (left) and large (right) markets

For the unstable markets the simulated fractions are very similar for both small and large groups (Figure B.10). 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.

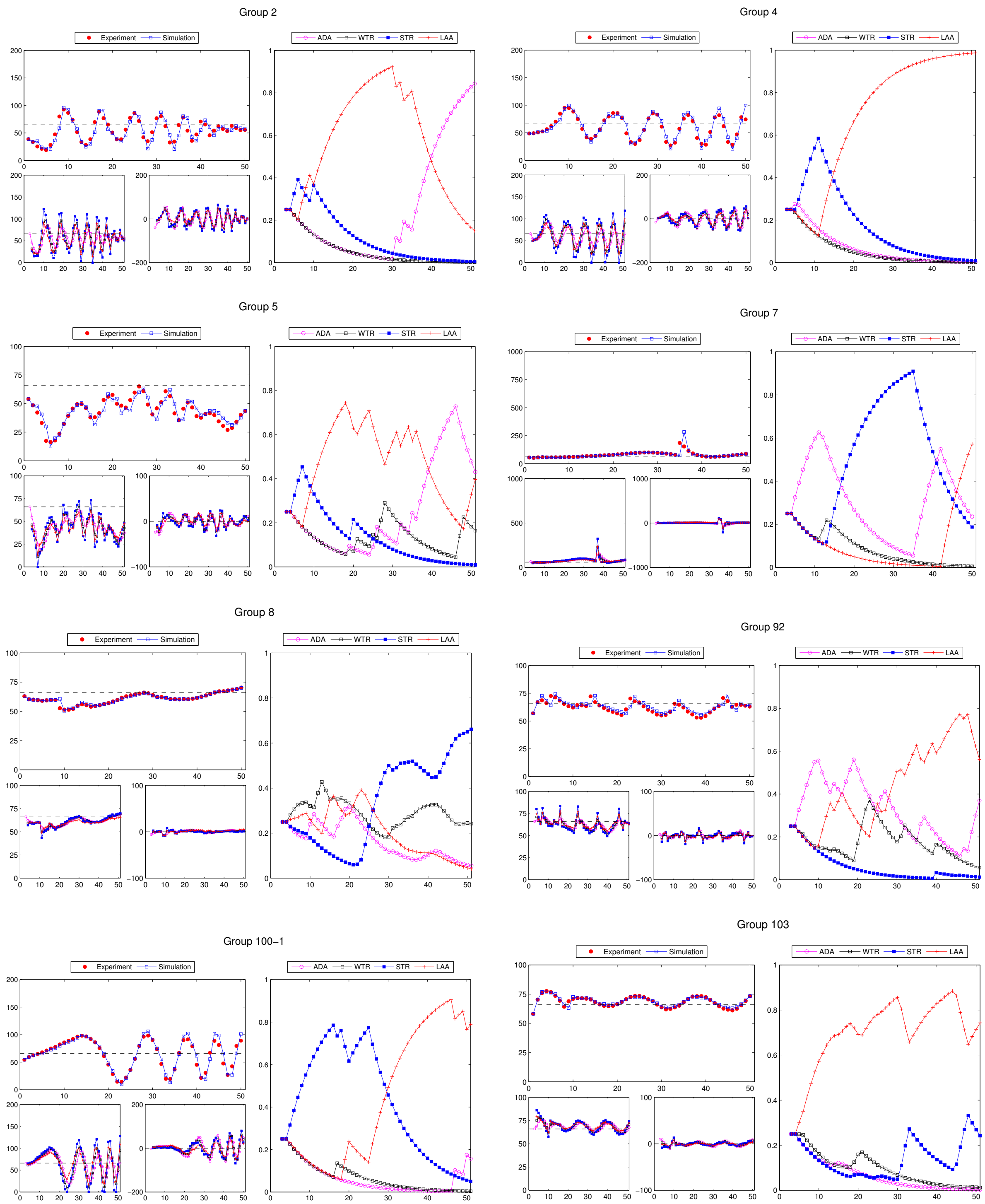


Figure B.11: HSM for stable markets in small (first 5 panels) and large (last 3 panels) groups.

Figures B.11-B.13 show the HSM simulated for the three different behavioural patterns 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 patterns

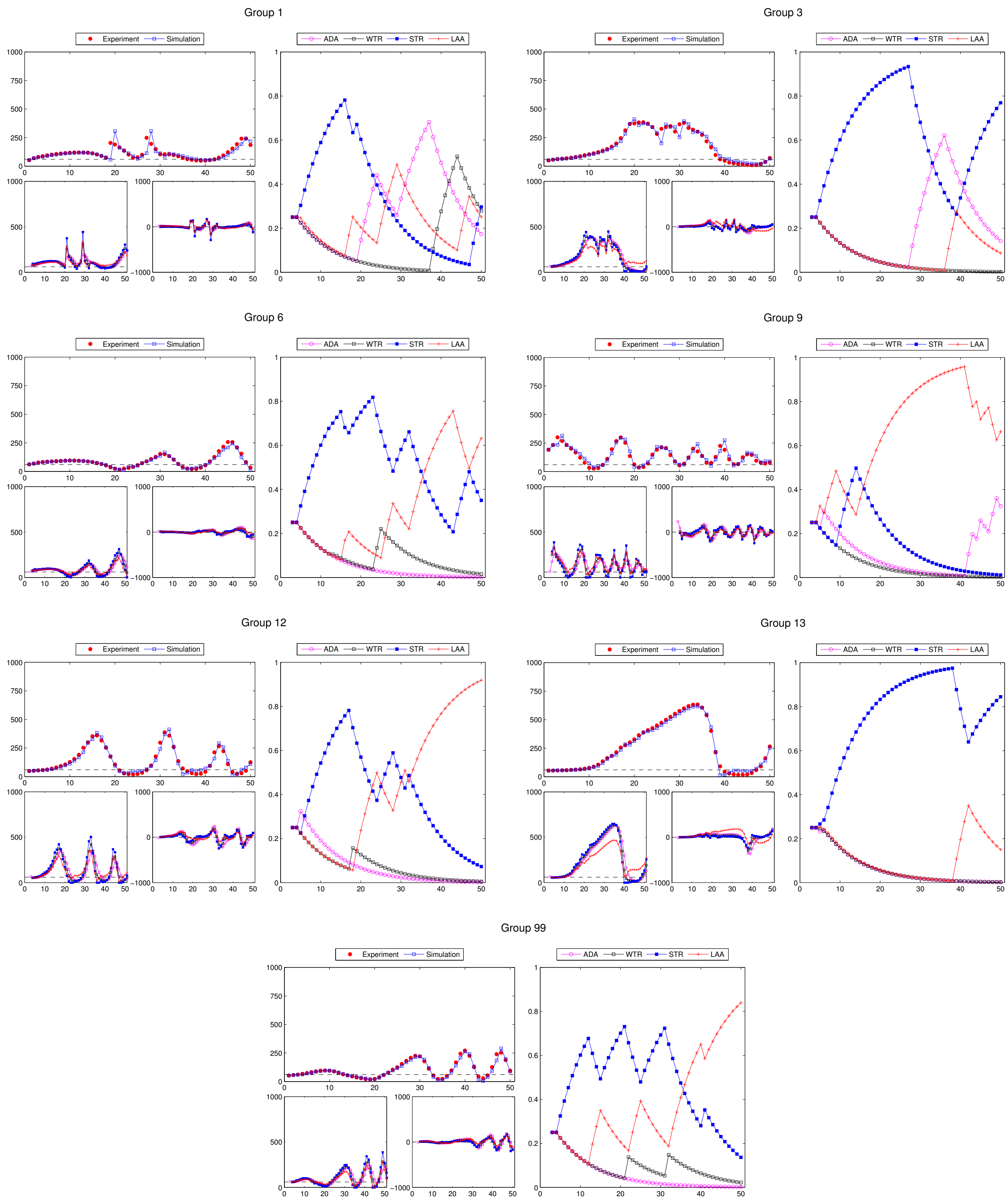


Figure B.12: HSM for large bubbles in small (first 6 panels) and large (last panel) groups.

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

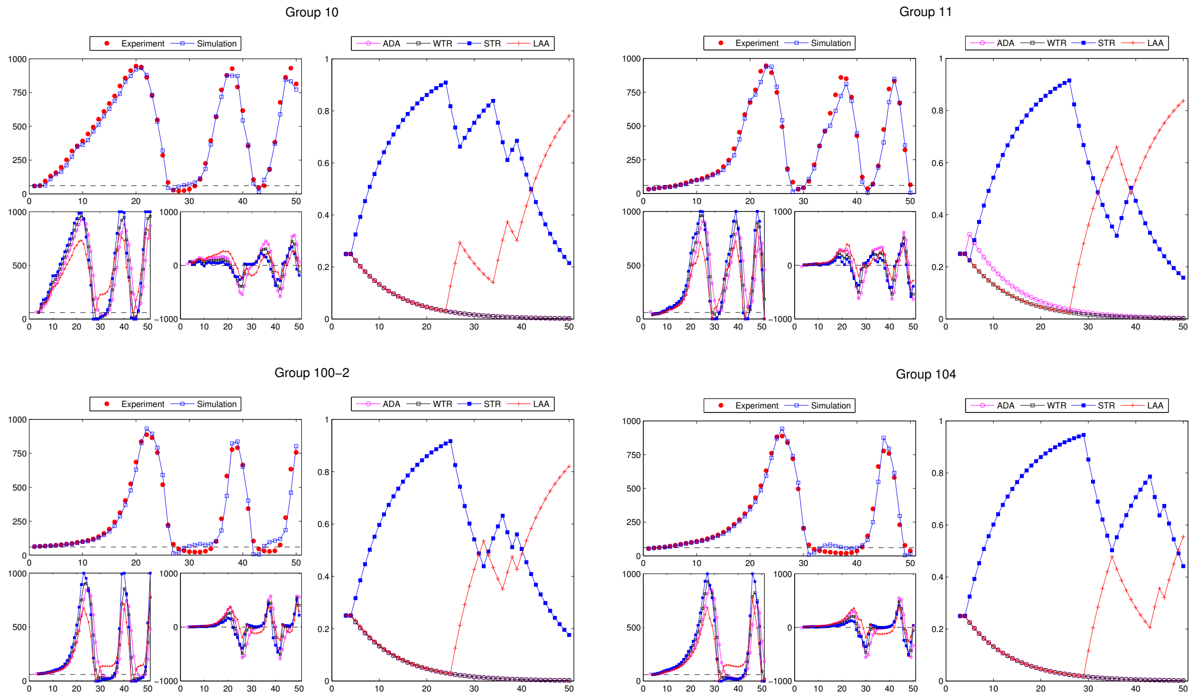


Figure B.13: HSM for very large bubbles in small (upper) and large (lower panels) groups.

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.

	Homogeneous rules						Heterogeneous rules			Fitted parameters		
							Fixed	Original	Fitted			
	Fund	Naïve	ADA	WTR	STR	LAA	fraction	HSM	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 B.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 B.2: Mean squared error for the different heuristics

Table B.2 lists the MSE for each market for each model, and the parameters for the best-fitting HSM-model.[7] 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,

[7]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.

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. (2019). Simple forecasting heuristics that make us smart: evidence from different market experiments. *Journal of the European Economic Association*, 17(5):1538–1584.

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.

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

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.

Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica*, 65(5):1059–1095.

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.

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.

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.