

Rational Expectations in an Experimental Asset Market with Shocks to Market Trends[☆]

Philipp Marquardt^{a,*}, Charles N Noussair^c, Martin Weber^b

^a*European Central Bank, Frankfurt, Germany*

^b*University of Mannheim, L9 1-2, 68131 Mannheim, Germany*

^c*University of Arizona, McClelland Hall 401X, Tucson, Arizona, USA*

Abstract

We construct an experimental asset market in which the time trend of the fundamental value is subject to a shock. The design of the experiment allows testing of whether prices adhere to Rational Expectations levels, and whether there is over- or under-reaction to new information. We find that prices conform closely to Rational Expectations and episodes of mispricing are rare. A meta-analysis allows us to update our beliefs about whether experimental asset markets exhibit a generic tendency to misprice, particularly in bearish environments.

Keywords: experimental asset markets, news reactions, price discovery, rational expectations

JEL codes: D84, G14, G40

1. Introduction

Expectations about the future are an integral element of capital markets, since individuals trade claims to uncertain future financial payoffs. Because traders are required to form expectations about the distribution of these future payoffs, they are faced with a difficult challenge: Since prices are simultaneously determined by expectations, traders are trying to hit a moving target. The elegant and general theoretical solution to this problem involves the assumption of *rational expectations*, which postulates that individuals form their expectations

[☆]We thank the German Research Foundation (DFG, Project ID: we993/15–1) for providing funding for this research. We also thank participants at the 2016 Experimental Finance Conference (Mannheim, Germany), the 2016 Economic Science Association North American Meetings (Tucson, Arizona, USA), the 2016 BEAM Workshops at Waseda University (Tokyo, Japan) and at the University Nice Sophia Antipolis (Nice, France), as well as the Research Seminar in Finance and Banking at Mannheim University, for helpful comments. We also thank John List for his comments on an earlier version of the paper.

*Corresponding Author

Email addresses: philipp.marquardt@mailbox.org (Philipp Marquardt),
cnoussair@email.arizona.edu (Charles N Noussair), weber@bank.bwl.uni-mannheim.de (Martin Weber)

according to the best estimates they have available of the distribution of exogenous random variables, and the application of the relevant economic theory (Muth, 1961).

For an asset that is finitely-lived, acquiring rational expectations boils down to forming beliefs about the risk-adjusted discounted sum of future dividends, and assuming that all traders in the market use this discounted sum as their limit price. As the rational expectations hypothesis has been the key assumption in many asset pricing models, its empirical foundation had for a long period of time remained largely unchallenged. This was because rational expectations are difficult to test, since both expectations and fundamental values are essentially unobservable. Parts of the economics and finance research communities began to become concerned, however, after Shiller (1981) observed that stock price volatility exceeds actual dividend fluctuations.

The emergence of research on experimental asset markets offered a new arena to test the rational expectations hypothesis, because it allowed researchers to have control over asset fundamentals, as well as to directly measure traders' expectations. Over the years, a standard repertoire of experimental designs for testing rational expectations and market efficiency has emerged. These were tailored to focus on specific functions of capital markets. Markets *disseminate* privately-held information, they *aggregate* dispersed information and they *establish a market value* for uncertain future cash flows. Early asset market experiments reinforced confidence in the efficiency of financial markets, because experimental results on the efficiency of markets for non-durable consumer goods (Smith, 1962) also seemed to extend to markets for assets with a life of two or three periods (Forsythe et al., 1982; Friedman et al., 1984). To study information dissemination, Plott and Sunder (1982) devised an experiment showing how insider information is incorporated into prices. Regarding information aggregation, Plott and Sunder (1988) documented how diverse information was aggregated into prices. Other experiments, such as Sunder (1992), provided support for even more complex dynamically rational behavior. He documented how subjects reduce their efforts in information acquisition, as markets become more informationally efficient, supporting the *information efficiency paradox* (Grossman, 1976; Grossman and Stiglitz, 1980). Finally, "learning-to-forecast" (LTF) experiments represented an important addition to the list of experimental asset market designs. In LTF experiments, the researcher maintains control over the price-setting mechanism, which allows the study of convergence to different expectational equilibria (e.g. Hommes et al., 2005).

Any experimental design inevitably must control for certain dimensions of the market environment in order to provide clean test beds for particular questions of interest. Thus, except for LTF experiments, the above-mentioned studies simplified the temporal dimension of financial markets and only studied assets with a life of one to three periods. Studying

long-lived instead of short-lived assets introduces important new issues: Do the long-term dynamics of asset prices and expectations adhere to underlying return generating processes, as suggested by rational expectations? Or might the opportunity to resell assets introduce speculative dynamics that cause prices to become detached from fundamentals?

The most influential experiment in the literature evaluating rational expectations in a market for a *long-lived* asset is that of Smith et al. (1988). Prior experiments had established the tendency of prices to converge to rational expectations when markets for assets with a life of 2 or 3 periods were repeated multiple times (Forsythe et al., 1982; Friedman et al., 1984) under identical conditions. However, Smith et al. (1988) studied the evolution of prices and beliefs over longer 15 - 30-period horizons, which allowed greater scope for the decoupling of prices from fundamentals to occur. Indeed, they observed that the rational expectations paradigm failed in their setting and the markets produced large bubbles and crashes.

The findings of Smith et al. (1988) have sparked numerous follow-up studies.¹ The bubble and crash pattern is quite robust, though some studies have identified mechanisms that abate bubble formation considerably. Mispricing typically decreases as subjects gain experience with repeated interaction in markets with an identical parametric structure (Dufwenberg et al., 2005). However, changes in parameters can cause bubbles to rekindle even when traders are experienced (Hussam et al., 2008). The process governing the time profile of the dividend and the terminal redemption value of the asset determine the fundamental value. The dynamic properties of the dividend process imply a time path of the fundamental value. The extent to which market prices adhere to fundamentals tends to vary depending on this time trajectory. When the fundamental value decreases over time, mispricing is greater than when it is increasing (Stöckl et al., 2014) or constant (Noussair et al., 2001). However, when the onset of the trend in fundamentals is delayed, prices are closer to fundamentals when the trend is decreasing than increasing (Breaban and Noussair, 2015).

While the emergence of bubbles in markets for long-lived assets has become somewhat of a stylized fact in the experimental research community, the evidence comes almost exclusively from studies that are based on the influential Smith et al. (1988) design. The applicability of the results from asset markets following this design to broader classes of environments has been questioned in recent years by a number of scholars (Lei et al., 2001; Oechssler, 2010; Kirchler et al., 2012). The suggestion is that the Smith et al. (1988) design with its deterministically declining fundamental value is especially prone to subject misunderstanding, and that this confusion is the source of its tendency to exhibit deviations from rational expectations pricing. The aim of this paper is to propose, construct, and study an

¹For comprehensive reviews see Palan (2013) or Powell and Shestakova (2016).

additional asset market design for studying price dynamics in long-lived asset markets. Our design includes elements that allow for testing the adequacy of different models of price and expectation formation.

In our design, the fundamental value of the asset that trades in the market is constant for two periods. Subsequently, there is a one-time shock to the trend in the fundamental value process, which can trend upward or downward thereafter. The market is designed to incorporate a number of special features. The first is that the predictions of Rational Expectations differ greatly from those of Myopic models. Because the shock occurs to the *trend* in the fundamental, forward-looking models that anticipate continuation of the trend can be easily distinguished from myopic models that consider only current dividend realizations but ignore the future trend. The second feature is that the onset of the trend is delayed rather than initiated at the outset of the market. The time interval before the trend allows the initial allocation of assets and cash to be reshuffled through trade, so that the prevailing conditions at the onset of the trend are endogenous rather than imposed by the experimenter. This can arguably make a difference, in that it means that at the onset of the trend relatively risk averse traders might hold fewer shares and more cash, and more capable traders may have increased their wealth by engaging in more profitable transactions than less capable traders. We consider this design feature to be important, since markets ought to presumably be in a risk-sharing equilibrium when value-relevant news shocks arrive. Otherwise, if the trend set in immediately upon the market open, some of the trade that would be observed in early periods would consist of portfolio rebalancing in reaction to the initial allocation, rather than in response to new information, or speculation regarding future prices.

Our environment can be briefly described as follows. Participants trade shares of a firm with a lifetime of 12 periods in a continuous double auction. The firm announces earnings each period, but does not pay out any interim dividends. Instead, it retains all earnings and is liquidated at the end of the life of the asset. At that time, all retained earnings are paid out to shareholders in the form of a one-time dividend per share. The earnings follow a two-point distribution that has zero expected value in the first two periods. After period two, however, the firm announces whether a regime with positive or negative expected value will prevail for the subsequent eight periods. This announcement induces a pronounced shift in the expected sum of earnings and thereby in the fundamental value of the firm. The setting therefore allows detection of under- and over-reaction to new information. The regime reverts back to an expected value of zero in the final two periods. At the beginning of each period, we also elicit subjects' beliefs about the average price that will prevail in that period. Subjects participate twice in the entire twelve-period market.

Two criticisms of the literature on experimental asset markets are that experimental subjects are often confused about the environment and that crucial features of asset markets in the field are left out of the experiment. We believe that the delayed onset of the fundamental value trend shock promotes understanding of the environment. Prior research has shown that a stationary fundamental value process is conducive to greater participant understanding (Kirchler et al., 2012) and we believe that the first two periods are helpful in promoting subject understanding. While we do not quiz subjects about their comprehension after the market has opened, the close adherence of prices and beliefs to Rational Expectations that we observe suggest that the environment was well understood. In our view, no experimental study could fully parallel the markets found in the field, nor should one try to do this, since it would involve adding too many complicating features to the environment. Our design does include an important feature that some other experimental environments have lacked, the arrival of news in a manner such that either over- or under-reaction can be clearly detected.

This setting allows us to test how well different models describe prices and beliefs in our markets. We propose and compare models assuming rational, trend-following and myopic expectations. Rational Expectations (RE) serve as our primary benchmark of interest since they are standard in finance. However, they have repeatedly been shown to fail to describe prices and beliefs in experimental studies employing extensions of the original Smith et al. (1988) design. Instead, prices and beliefs tend to adhere to a mixture of adaptive and trend-following models (Haruvy et al., 2007; Carle et al., 2016). Therefore, we specify a Trend-following and a Myopic model as alternatives. Assuming myopic agents is not implausible here. In our setting, the firm's earnings accumulate until they reach the final liquidation value. These accumulated earnings are a natural, though incorrect, measure of share value that might seem plausible to traders who do not take the future into account. Experimental results on myopic loss aversion (Thaler et al., 1997) and the stock market's tendency to overweight near-term earnings (Abarbanell and Bernard, 2000) suggests that some subjects might indeed use a myopic model that is based on the current stock of retained earnings.

Finally, our design provides a test environment for enriching the continuing debate on the extent to which asset prices underreact or overreact to new information. A variety of behavioral finance models and empirical studies has emerged to challenge Fama's conjecture that stock price underreaction to news is about as common as overreaction and most likely the source of misspecified asset pricing models (Fama, 1998). Concerns about such *joint hypothesis testing* arise frequently in event studies that analyze the post-earnings announcement drift or earnings momentum. These studies typically analyze share price reactions to corporate announcements as incidents for testing market efficiency (Fama et al., 1969; Bernard and Thomas, 1989; Jegadeesh and Titman, 1993). Event studies are even employed

as evidence in court for securities fraud litigation (Brav et al., 2014). Thus, the question of whether markets react efficiently to news or not has immediate financial impact. We devised our experimental set-up to mimic the research design in these event studies. We thereby complement the event-study literature with experimental evidence.²

Our results exhibit a number of consistent patterns. We do not observe pronounced bubbles and crashes. Instead, prices in our markets adhere closely to the levels predicted by the Rational Expectations Model. After traders gain some experience, the Rational Expectations Model outperforms the two alternative models in predicting prices. Its performance relative to the Trend Model, which assumes the continuation of current price trends, among inexperienced participants depends on the sign of the news shock. The RE Model predicts prices more accurately than the Trend Model in markets with a negative shock, but the Trend Model performs marginally better in markets that experience a positive shock. In general, however, market prices on average incorporate news shocks rapidly and there are few signs of systematic under- or overreaction. With respect to beliefs, the Rational Expectations and the Trend models perform comparably well in the first market in which traders participate, while the Myopic Model fits our data less well than the other two models. Prices and beliefs move even closer to Rational Expectations levels when subjects interact in a second market after they have gained some experience with the market. As both the Rational Expectations and the Trend models both have some power to explain the data, we introduce and estimate a hybrid model in Section 6. This model allows individuals' expectations to adapt and to weight the different models based upon relative predictive performance. The hybrid model has better explanatory power than any of the three models alone, but also demonstrates that subjects' beliefs converge rapidly towards Rational Expectations over time.

In addition to testing the relative fit of the different models, we analyze whether subjects form their expectations differently based on their cognitive ability, as measured by the cognitive reflection test (CRT) of Toplak et al. (2014). We find that individuals with higher CRT scores have beliefs that exhibit closer adherence to rational expectations and less conformity to previous price trends. While subjects with higher CRT-scores adhere more closely to rational expectations, they only marginally predict prices better, which indicates that other capabilities are also relevant for anticipating price dynamics (Bruguier et al., 2010). Given the little confusion our design appears to create among subjects, we do not find a significant

²Our study is related to a number of experiments that study the impact of different fundamental value time paths on mispricing, and the prevalence of under- or overreaction. For instance, Lin and Rassenti (2012), Kirchler (2009), and Stöckl et al. (2014) measure the adherence of prices to fundamental values under different random walk regimes. These designs differ from ours in that subjects do not experience a large, salient shock to fundamentals as they do in our experiment, and beliefs are not elicited. We believe that it is this large shock to the trend that makes our design well-suited to test for rational expectations.

relationship between a cohort’s average CRT score and mispricing in our markets, as other studies do (Breaban and Noussair, 2015; Bosch-Rosa et al., 2018).

Finally, we conduct a meta-study to compare the extent of price efficiency in our setting with related experimental designs of long-term asset markets. Maniadis et al. (2014) propose a Bayesian framework to evaluate the implication of new experimental findings for established research areas. By combining the research prior and our statistical power, we demonstrate that prior beliefs about the extent of mispricing in experimental markets should be revised down considerably in light of our findings. The results are reported in Section 5.

2. Theory and hypothesis development

The asset in our experiment has a lifetime of $T = 12$ periods, and is described as representing a share of a company. In each period $t = 1, \dots, T$, the company retains earnings per share of e_t , which are added to the initial-level retained earnings e_0 and the earnings accumulated until period t . In period T , the firm distributes a dividend to the holder of each share. The dividend paid on each share is equal to $D_T = \sum_{t=0}^T e_t$.

The experiment is designed to allow the application of several competing models that make point predictions for both prices and trader expectations in period t . Let p_t denote the price in period t , and let b_t be the belief of traders at the beginning of period t , about the price that will prevail in period t . For the purpose of computing our benchmarks, expectations are assumed to be homogeneous for all subjects. One benchmark is the *Rational Expectations Model*, abbreviated henceforth to RE. Under this model, prices in period t , as well as beliefs submitted at the outset of period t , are the expectation of the final dividend conditional on the retained earnings realizations prior to period t . That is:

$$p_t^{RE} = b_t^{RE} = \mathbb{E}_t(D_T | e_0, \dots, e_{t-1}) \quad (1)$$

Prior research on experimental asset markets has noted, however, that many participants in experiments fail to anticipate the future evolution of the value of an asset, even when it can be deduced from information that is currently available (Smith et al., 1988). We thus consider a boundedly rational alternative model, called the *Myopic Model*, in which individuals believe that the value of the asset is the sum of the retained earnings accrued until the present. Prices depend only on the past, and there is no forward-looking element in the model. Prices and beliefs under the Myopic Model are given by:

$$p_t^{ME} = b_t^{ME} = \sum_{s=0}^{t-1} e_s \quad (2)$$

Other previous experimental work has shown that some traders' expectations of future prices are consistent with the extrapolation of previous price trends (Haruvy et al., 2007; Carle et al., 2016). We capture this with the *Trend Model*, in which prices and expectations satisfy:

$$p_t^{TR} = b_t^{TR} = p_{t-1} + (p_{t-1} - p_{t-2}) \quad (3)$$

The Trend Model is also exclusively backward looking, but uses observed prices rather than earnings information as input. While we presume that models containing free parameters, such as adaptive or heuristic switching models (Hommes et al., 2005), might provide a better fit to the data, we choose the three models above to apply to our data, because they have no free parameters and make precise unambiguous point predictions of prices and beliefs in each period after period 1. They are thus on an equal footing with each other in this regard. The predictions of the Rational Expectations and the Myopic models for price and belief trajectories in our experimental environment are shown in Figure 1. For an example of a price/belief path predicted by the Trend Model, which uses actual price data and thus depends on the history of activity in the market, see Figure 4 in Section 4.2.

To compare the models with regard to how well they predict prices, we measure the root mean squared error (RMSE) between actual observed market prices and the models' predictions at the market level. For example, for the RE Model, we compute:

$$RMSE(p_t) = \sqrt{\frac{\sum_3^T (p_t^{RE} - p_t)^2}{T}} \quad (4)$$

The RMSE is calculated for the other models as well. We calculate the RMSEs for each market and the model with the lowest average RMSE is then considered as the most accurate. We estimate the models' fit using observations beginning in period 3, which is the time of the onset of the positive or negative trend shock.

For beliefs, we compare the root mean squared error between the model prediction and observed individual forecasts for all markets. For the RE Model, this implies:

$$RMSE(b_t) = \sqrt{\frac{\sum_i \sum_3^T (b_t^{RE} - b_{it})^2}{T * I}} \quad (5)$$

where $i = 1, \dots, I$ indexes the subjects within a period. The fit of the two remaining models are measured similarly, by using each model's predicted values.

In addition to measuring the models' fit to the belief and price data, we evaluate several hypotheses in the next subsection. The hypotheses originate in previous experimental research and concern the presence of specific patterns in the accuracy of the models. To

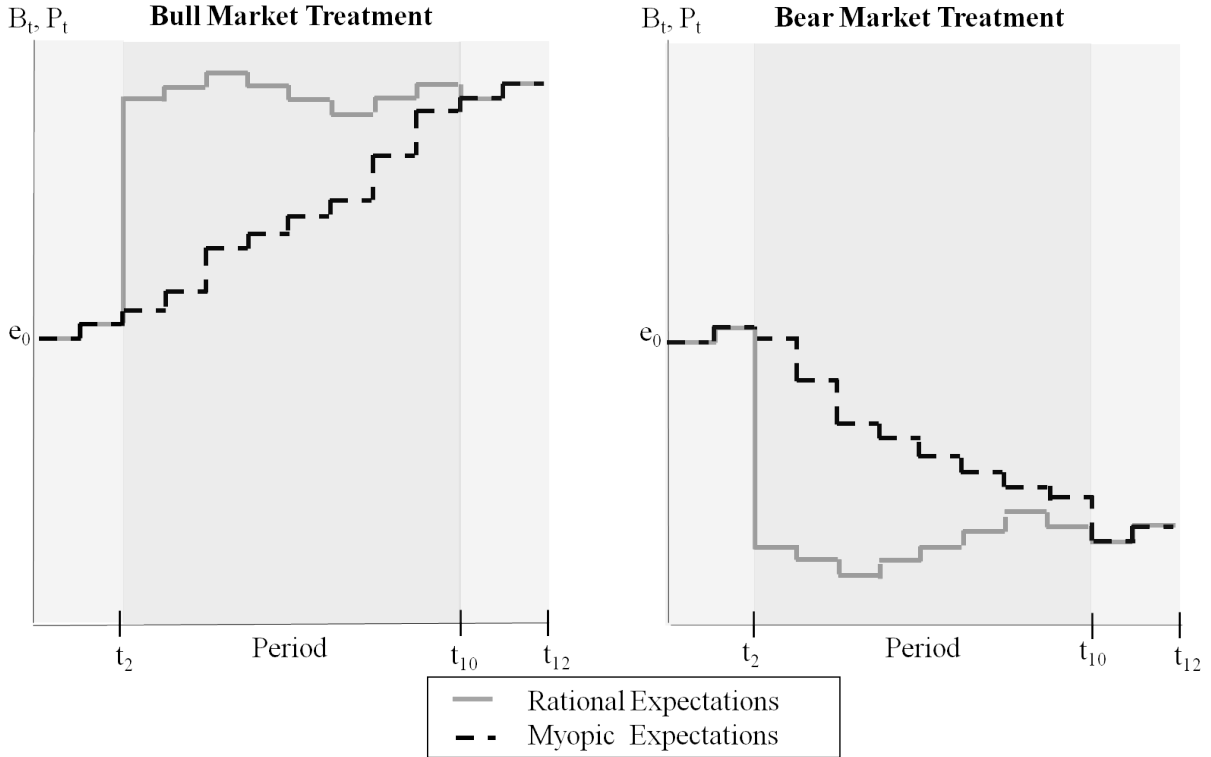


Fig. 1. Price and belief trajectories under the RE and the Myopic models

This figure compares the predicted price and belief trajectories under the Rational Expectations Model (solid gray line) and the Myopic Model (dashed black line). The models make distinct predictions of prices following the shift to a positive or negative trend after period 2. This shock to expectations allows for a direct test of the Rational Expectations paradigm. The panel on the left depicts a typical time profile of the two models in the Bull Market treatment, in which traders are informed before period 3 that a positive trend in the dividend process will be in effect until period 10 (inclusive). The panel on the right corresponds to the Bear Market treatment, in which there is a negative trend in effect from periods 3 - 10.

consider these hypotheses, we compute several measures of bubble magnitude that are commonly used in experimental asset markets. As indicated earlier, we consider markets to be mispriced, whenever prices deviate from the levels implied by rational expectations. Our main measure of mispricing is the relative absolute deviation (RAD) introduced by Stöckl et al. (2010). It is computed by taking the average absolute difference between mean period transaction prices and RE levels across periods and normalizing it by a market's average rational expectations value:

$$RAD = \frac{1}{T} \sum_{t=1}^T \frac{|\bar{P}_t - RE_t|}{RE} \quad (6)$$

A RAD of 10 percent means that on average, prices per period deviate by 10 percent

from the RE level over the life of the asset. In addition to our primary measure RAD, we also report the Relative Deviation (RD):

$$RD = \frac{1}{T} \sum_{t=1}^T \frac{\overline{P}_t - RE_t}{\overline{RE}} \quad (7)$$

RD provides us with a measure of under- or overpricing relative to RE levels. Additionally, when calculated at the period-level, RD allows us to test for patterns of under- or overreaction to the earnings shocks in our experiment. Finally, we also report Geometric Deviation (GD) (Powell, 2016), Geometric Average Deviation (GAD), Price Dispersion (King et al., 1993) and Turnover to ensure robustness and comparability to earlier studies. We use RAD, RD, as well as measures of belief accuracy, belief dispersion, and earnings, to evaluate a number of hypotheses. These hypotheses originate in patterns that have been documented in previously conducted experiments in related environments. The hypotheses can be grouped into two categories, based on whether they describe market-level or individual-level behavior. The first two hypotheses that we advance concern market behavior, and specifically the efficiency of the market.

Determinants of price efficiency. The Rational Expectations Model is consistent with the presence of fully rational agents and common knowledge that traders are rational. Models with these assumptions exhibit an increasing tendency to perform well in the Smith et al. (1988) environment as individuals repeat the task (Smith et al., 1988; Haruvy et al., 2007). Mispricing can reemerge, however, even as individuals accumulate experience, when some parameters of the asset market are changed from one repetition of the market to the next (Hussam et al., 2008). In our experiment, we change the parameters from the first to the second market, and it is unclear from the previous literature whether we would expect an improvement in the accuracy of the RE Model in the second market. Subjects in our setting cannot necessarily extrapolate their experience from the first to the second market, because they participate in both Bull and Bear markets and because learning processes have been shown to differ between the gain and the loss domains (Kuhnen, 2015). Thus, we hypothesize that the model’s accuracy would not change over time.

Hypothesis 1. *Prices are equally close to the Rational Expectations Model prediction in the first and the second market.*

The second hypothesis concerns how the accuracy of the RE model compares under the two price trajectories we study. The hypothesis also originates in previous experimental work. In the SSW-paradigm, deviations from fundamentals are less pronounced when the

fundamental value is increasing over time rather than decreasing (Stöckl et al., 2014; Kirchler, 2009), but the result is reversed when the trend sets in after a delay (Breaban and Noussair, 2015) as it does in our experiment. Thus, because of the mixed previous results, we hypothesize that adherence to rational expectations is equally likely under the two fundamental value regimes.

Hypothesis 2. *The magnitudes of deviations from the RE model are similar in Bull and Bear markets.*

Under- vs. overreaction to news. The first two hypotheses concerned the general extent of mispricing across treatments, but were silent about whether prices would underreact or overreact in response to the shock after period 2. To formulate a hypothesis about the pattern of price reactions, we consider short-and long-term reactions in tandem. This is because prominent behavioral finance models that offer an alternative to the efficient market hypothesis provide a unified account of price reactions for both the short and the long term. Therefore, they can only be tested meaningfully by considering both time horizons jointly.³ We outline two prominent models that incorporate behavioral and cognitive biases and make contrasting predictions about price patterns: the model by Barberis et al. (1998) - henceforth BSV - and by Frazzini (2006). Since all information is public in our setting, we do not consider popular models such as Daniel et al. (1998) or Hong and Stein (1999), as they study the interplay of public and private news. BSV predict initial underreaction to news, followed by subsequent overreaction and reversal. Underreaction is driven by *conservatism bias* that makes agents reluctant to revise previously held views and overreaction is driven by *representativeness bias* that leads agents to extrapolate price patterns from small samples. Contrastingly, Frazzini (2006) predicts that prices underreact to news due to the *disposition effect* both in the long and the short term. After positive news, investors try to lock-in their gains and depress prices, whereas after negative news investors are reluctant to sell and provide price support. This mechanism slows convergence of prices towards new fundamental levels.

Fama (1998), who is skeptical about such behavioral finance models, provides a natural third alternative. According to him, empirical evidence suggests that overreaction in the

³Another reason is that the debate about the efficiency of capital markets arises primarily when looking at the cross-section of *long-term* stock returns. Studies using short event-windows around corporate announcements find that stock prices typically adjust rapidly to new information and continue to fluctuate randomly thereafter. However, other studies that include longer time horizons following the event document long-term under- or overreaction. Fama (1998) criticizes attempts to reject market efficiency in response to these findings for two reasons. First, there is an even split between studies documenting under- and overreaction. Second, calculating long-term excess returns is subject to measurement issues, which are less severe in short-term event studies.

long-term occurs just as frequently as underreaction. Furthermore, he argues that these models are tailored to specific anomalies and circumstances, but fail to beat the efficient market hypothesis on the whole. Since our aim was to study price discovery in a novel experimental design, our set-up is not intended as a direct test of any of the two behavioral models. In contrast, Weber and Welfens (2007), for instance, conducted a direct test of Frazzini (2006), as well as of Grinblatt and Han (2005), by deliberately creating subject groups with different susceptibility to the disposition effect. Our design would be ideal for testing models relying on the disposition effect, since we shock the fundamental value *trend*, which induces the salient reference point required in those kind of models. However, in this paper, we do not employ such exogenous variation in susceptibility to the disposition effect because we are more interested in overall price dynamics. Therefore, we take Fama's rejection of systematic under- or overreaction as our Null hypothesis.

Hypothesis 3. *Underreaction to the news shock is as common as overreaction in our markets. Relative Deviation (RD) is not systematically different from zero.*

Individual behavior. We also advance three hypotheses with regard to individual behavior. These are also based on evidence from prior studies. The first relies on the fact that market activity tends to stabilize as a market is repeated and is stated below. In spite of the fact that subjects have homogeneous information about the asset's fundamental value process, previous experimental work has shown individual beliefs about prices to be heterogeneous. This heterogeneity tends to decrease with repetition of the market (Carle et al., 2016), though this evidence comes from markets that are repeated with stationary parameters. If the common experience that individuals have in the first market affects beliefs in the second market, we would expect beliefs to be more homogeneous in the second market than the first. Because we have already readily hypothesized that pricing relative to rational expectations is similar in Markets 1 and 2, we would also expect that beliefs would exhibit similar adherence to rational expectations in market 1 and market 2.

Hypothesis 4. *Expectations are more homogeneous in the second than in the first market. They differ similarly from rational expectations levels.*

The remaining two hypotheses concern the relationship between individual trader characteristics and outcomes, both at the individual and the market levels, and can be found in Appendix A. They serve to further reconcile our study with previous experimental findings.

3. Experimental design

3.1. General features

The experiment consisted of 16 experimental sessions conducted at the University of Mannheim. In each session, nine subjects participated in two sequential asset markets. Each market consisted of twelve trading periods, during which subjects could use cash to trade shares in a computerized continuous double auction. Dividends, earnings, and cash holdings were denominated in terms of an experimental currency, called Taler. Trading was conducted in terms of Taler. The conversion rate of Taler to Euro was 500 Taler = 1 Euro (1 Euro = \$US 1.20 approximately at the time the experiment was conducted).

There were two treatments, Bull and Bear, reflecting the time trend of the fundamental value. The protocol followed a within-subject design in which each cohort of traders participated in one Bull and one Bear market. The order of treatments was counterbalanced, so that Bull markets preceded Bear markets in some sessions and vice versa in other sessions.⁴ Subjects were recruited using *ORSEE* (Greiner, 2003). The market was implemented using an adapted version of the standardized software *GIMS* (Palan, 2015) on the platform *z-tree* (Fischbacher, 2007).

3.2. Asset value and information structure

The asset was described as a share in a firm. It was indicated that the firm did not pay any interim dividends, but instead announced per-period earnings at the end of each of the twelve periods. These earnings were retained and the accumulated earnings were distributed to the shareholders at the end of period 12. The earnings process followed a random walk and could be subject to different regimes, which depended on the firm's business conditions. These business conditions varied over time according to the following scheme, which was common knowledge to all subjects:

Period 1 - 2: Business conditions are neutral. Period earnings are +10 Taler or -10 Taler with equal probability.

Period 3 - 10: If business conditions are *good*, period earnings are +30 Taler or +10 Taler with equal probability. If business conditions are *bad*, period earnings are -30 Taler or -10 Taler with equal probability.

⁴Two sessions were exceptions to this. In one of these, two Bull markets were conducted and in the other two Bear markets were run. See Table 12 in Appendix B.

Period 11 - 12: Business conditions are neutral. Period earnings are +10 Taler or -10 Taler with equal probability.

Accordingly, at the beginning of period 1, it was announced that business conditions would be neutral for the first two periods.⁵ At the beginning of period 3, it was announced publicly that business conditions were either good or bad. At this time, it was indicated that business conditions would remain the same until the end of period 10 and return to neutral at the beginning of period 11. At the beginning of period 11, subjects were reminded that conditions had returned to neutral. The instructions read before the experiment began also indicated that business conditions would be neutral in periods 1 and 2, either good or bad for periods 3 - 10, and neutral in periods 11 - 12.

Since the fundamental value of the share was determined by the expected future earnings realizations, it crucially depended on this one-time announcement of business conditions at the outset of period 3, and less so on the earnings announcements in each of the other 11 periods. This is because the period 3 announcement determined the expected value of the next 8 earnings announcements. Because this expected values was strictly positive or negative, it had a large impact on the final dividend. Figure 2 illustrates the earnings process.

3.3. Elicitation of beliefs

At the beginning of each of the twelve trading periods, before the market opened, but after any announcement of business conditions, subjects were asked to make a forecast of the upcoming period's average transaction price. The average transaction price was weighted by the quantity of units traded in each transaction. Subjects received 20 cents for each prediction that was within 10 percent of the actual average price. They also had to indicate how sure they were about their prediction on a five-point scale. Earnings from predictions were added to take-home payments of subjects, but were unavailable to be spent on assets in the market.

3.4. Timing of activity in a session

At the beginning of each session, subjects followed a detailed visual presentation giving an overview of all of the phases of the experiment, including the instructions for the market

⁵The choice of a two-period, rather than a longer, interval before a trend set in reflected a trade-off between (1) allowing time for individuals to rebalance their portfolios, (2) making sure that the trend was in effect long enough so that the Myopic and RE models made very different point predictions, and (3) ensuring that the entire market was short enough so that two markets could be completed in each session, and the effect of experience could be studied.

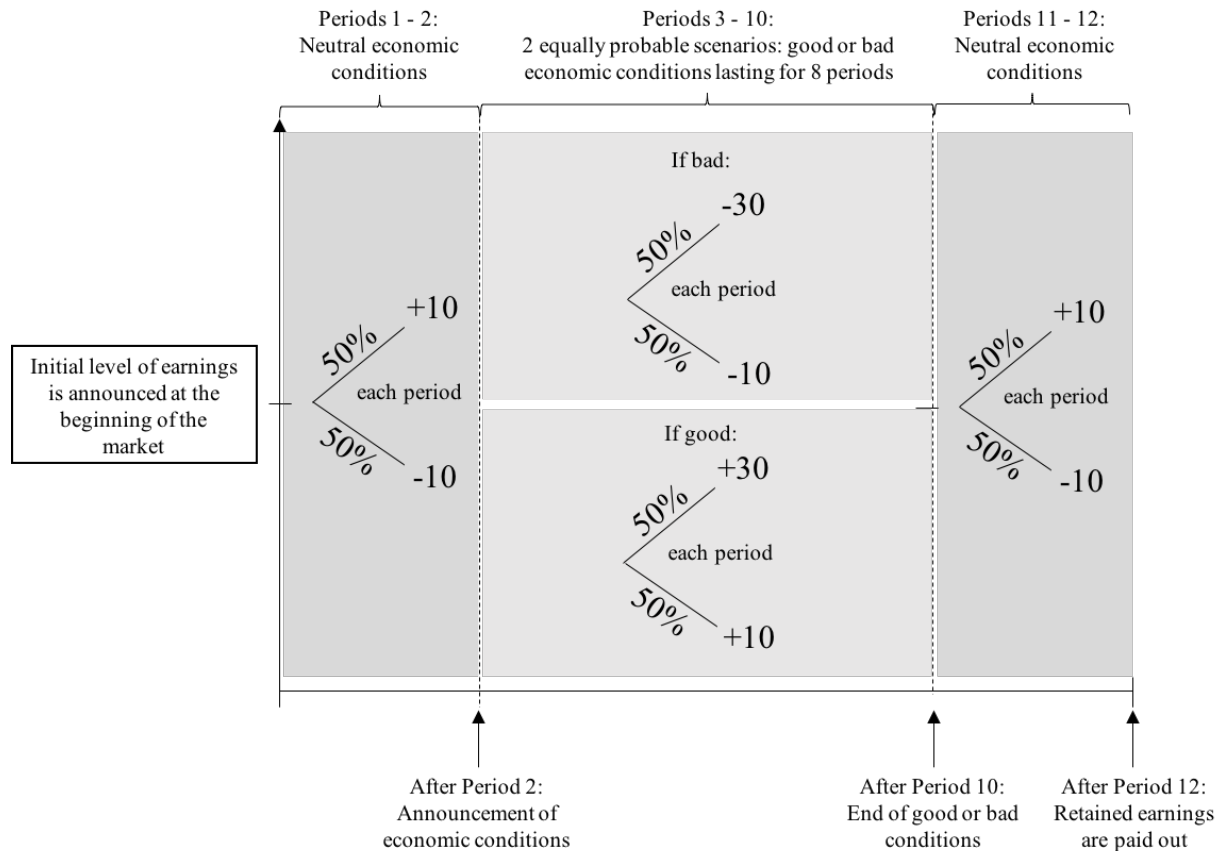


Fig. 2. Time Profile of the Asset's Fundamental Value

This figure illustrates the earnings processes and timing of events that determine the fundamental trajectory of the asset in our experiment. The *good* state prevailed in the Bull treatment, while the *bad* state was present in the Bear treatment. The experimental instructions contained the same figure in German language. Values are displayed on a per-share basis.

experiment, and the fact that they would be asked to take additional quizzes. They were allowed to ask questions during the presentation. In this presentation, the information structure of the market was presented publicly and in great length to ensure common knowledge among participants. Next, subjects played three practice periods of the market, which did not count toward participants' earnings, to make them familiar with the trading platform and sequence of events. In these practice periods, the parameters were chosen to be very different in magnitude from those in effect during the rounds that counted. This was to ensure that there were no carryover effects from the practice rounds to later rounds. During these practice periods, subjects could again ask questions.

After the completion of these practice periods, we administered the extended, seven-item, cognitive reflection test used by Toplak et al. (2014) to measure cognitive ability. Afterwards,

risk preferences of orders 2 - 4 (risk aversion, prudence and temperance), were elicited using the task of Noussair et al. (2014).⁶ Subsequently, and before the market opened, subjects had to correctly answer a number of comprehension questions about the asset market, specifically concerning the information structure and earnings process, before being allowed to continue. The market then opened and two twelve-period markets were conducted.

Each trading period was 2 minutes in duration. Trading was conducted with continuous double auction rules (Smith, 1962) and implemented with the z-tree platform (Fischbacher, 2007) and the GIMS market platform (Palan, 2015). At the end of the period, the firm would announce its earnings for the period. This sequence would repeat until period 12. In addition, at the end of period 2, the business conditions for the 8 following periods were announced. After each trading period, subjects would see an overview screen indicating the number of shares and the amount of cash that they currently possessed. It also indicated the total retained earnings per share at the end of each period, and the average market price in each period up to the present. The screen did not indicate the market value of a participant's current position or the expected final dividend payout based on the information currently available. Thus, subjects had to derive expected payouts, which stands in contrast to earlier experiments that readily provided subjects with these values (Smith et al., 1988).

Subjects' earnings consisted of several components. Firstly, they were rewarded for correct answers to the cognitive reflection test. If they answered all of the questions correctly, they received 1 Euro. Secondly, in the risk-attitude elicitation task, one of the lotteries was chosen at random to be paid. Thirdly, their end-of-period wealth in one of the markets was converted into Euros and theirs to keep. The market was chosen randomly and a different draw was made for each subject. If an individual ended the market with a short position, the dividend she owed would be subtracted from her cash holdings at the end of the market. Fourthly, in the market that counted, subjects received 0.20 Euros for each forecast that was within 10 percent of the actual average transaction price. The incentivization of forecasts was chosen to be sufficiently low to discourage manipulation of the market by a trader in order to make her predictions more accurate. Earnings for the session averaged 18 euros and each session took about two hours.⁷

⁶Sessions 1 and 2 employed the risk preference elicitation protocol of Holt and Laury (2002) and the original cognitive reflection test by Frederick (2005), which uses a three-item scale. Due to our desire to achieve a larger dispersion in outcomes, we switched to the list of questions employed in Toplak et al. (2014). Therefore, in the results section, when using the CRT in our analysis, the first two sessions are omitted. Also, beginning in session 3, we employed the Noussair et al. (2014) protocol to measure risk attitudes.

⁷While the sessions in our experiment are longer than in many other studies, they are shorter than in a number of experiments in the experimental asset market literature, including some that some of us have conducted. For example, the experiment reported in Haruvy et al. (2007) was three hours in duration, and that described in Dufwenberg et al. (2005) took 2 1/2 hours. The sessions reported in Lahav (2011) were

3.5. Market parameters

The parameters in effect in the markets are summarized in Table 1. Each session included three types of traders, differing in their initial level of endowment of cash and shares. The first two sessions had an initial cash-to-asset ratio⁸ close to 1 and short-selling was prohibited. Sessions 3 - 16 had short-selling enabled and a higher cash-to-asset ratio of 2. Agents could hold a short position equal to 12 units, a quantity equal to the highest initial endowment.⁹ These changes to the initial protocol were made, beginning in session three, to avoid mispricing or under-reaction due to limits to arbitrage, either because of a shortage of cash or binding short selling constraints. The earnings realizations for good and bad regimes were created by a random number generator. We used three different sequences of random draws, one each for sessions 1-5, 6-8 and 9-16, respectively.

Table 1: Market parameters across different sessions

	Sessions			
	1 - 2	3 - 8	9 - 16	
Endowments (Cash / Shares)			Bull	Bear
Trader type I	500 / 12	1000 / 12	1000 / 12	6000 / 12
Trader type II	2000 / 7	4000 / 7	4000 / 7	9000 / 7
Trader type III	3500 / 2	7000 / 2	7000 / 2	12000 / 2
Initial FV	300	300	300	620
Initial Cash-to-Asset-Ratio	1:1	2:1	2:1	2:1
Short sales	prohibited	allowed	allowed	
CRT protocol	3 item	7 item	7 item	
Risk preference measures	HL (2002)	NTV (2014)	NTV (2014)	
Average earnings	16 Euros	19 Euros	21 Euros	22 Euros

also over 3 hours in length. It is also the case that the second market is pricing more closely to the fully rational model of Rational Expectations than the first, suggesting that fatigue is not playing a role in these sessions; otherwise we would presumably observe a decline of rationality near the end of the session.

⁸The cash-to-asset ratio (Caginalp et al., 2008) in a market is equal to the total amount of cash available to traders, divided by the fundamental value times the total stock of units in the market. It can be thought of as the ratio of the largest long position to the largest short position that a representative individual can take (Noussair and Tucker, 2016), in the absence of the ability to borrow and to effect short sales.

⁹In the Smith et al. (1988) paradigm, the presence of short selling possibilities does not induce rational expectations. Instead, it lowers prices more generally, and the greater the short-sale capacity, the lower prices become (Haruvy and Noussair, 2006).

4. Experimental results

4.1. Summary of the data

The time series of average period transaction prices in each of the 32 markets is shown in Figure 3. The figure also shows the RE price trajectory as lines and the beliefs that individuals submit in the current period as points. Figure 3 reveals a number of broad patterns. First, prices tend to be closer to RE levels in the second market that a group participates in than in the first market. Second, prices are very close to RE levels late in the life of an asset. Third, the dispersion of beliefs tends to decrease within a session. On the whole, prices and beliefs seem to adhere relatively closely to rational expectations levels, especially in the second market. On rare occasions, we also observe that bubbles form or that prices in the first period start substantially below the firm's initial earnings level and slowly gravitate toward the rational expectations level. In these instances, prices and beliefs tend to follow previous price trends. However, such patterns seem to be confined to Bull markets and first markets and do not represent the norm in our study.

4.2. Model comparison

We now compare the accuracy of the three models in describing behavior. A visual impression is conveyed by Figure 4, which displays the models' fit to median prices and median beliefs across our treatments. A number of qualitative observations emerge from the figure. The Myopic Model achieves the lowest accuracy in describing both prices and beliefs. Thus, subjects do not simply take the accumulated earnings in a given period as a basis for forming their beliefs. Instead, they seem to use a mixture of the Trend and the RE Models to form their beliefs. This behavior is reflected in prices. To support this visual impression, we calculate the root mean squared error (RMSE) of all models in our sample. The results for prices and beliefs are shown in Table 2.

The models are also compared for the Bear and Bull markets, as well as for the first and second markets, separately. We first compute the RMSE of a particular model individually for each market and then pool across different treatment combinations. Thus, the values in each cell refer to the average RMSE from pooling 16 markets. The three columns on the right test whether the RMSEs are significantly different from one another. The data suggest that price dynamics differ across treatments, since the fit of our proposed models varies between Bull and Bear Markets. In Bull markets, prices tend to be better described by the Trend Model (RMSE = 79.52) than the RE Model (RMSE = 86.62). Yet, the difference in accuracy between both models is not statistically significant. In Bear markets, however,

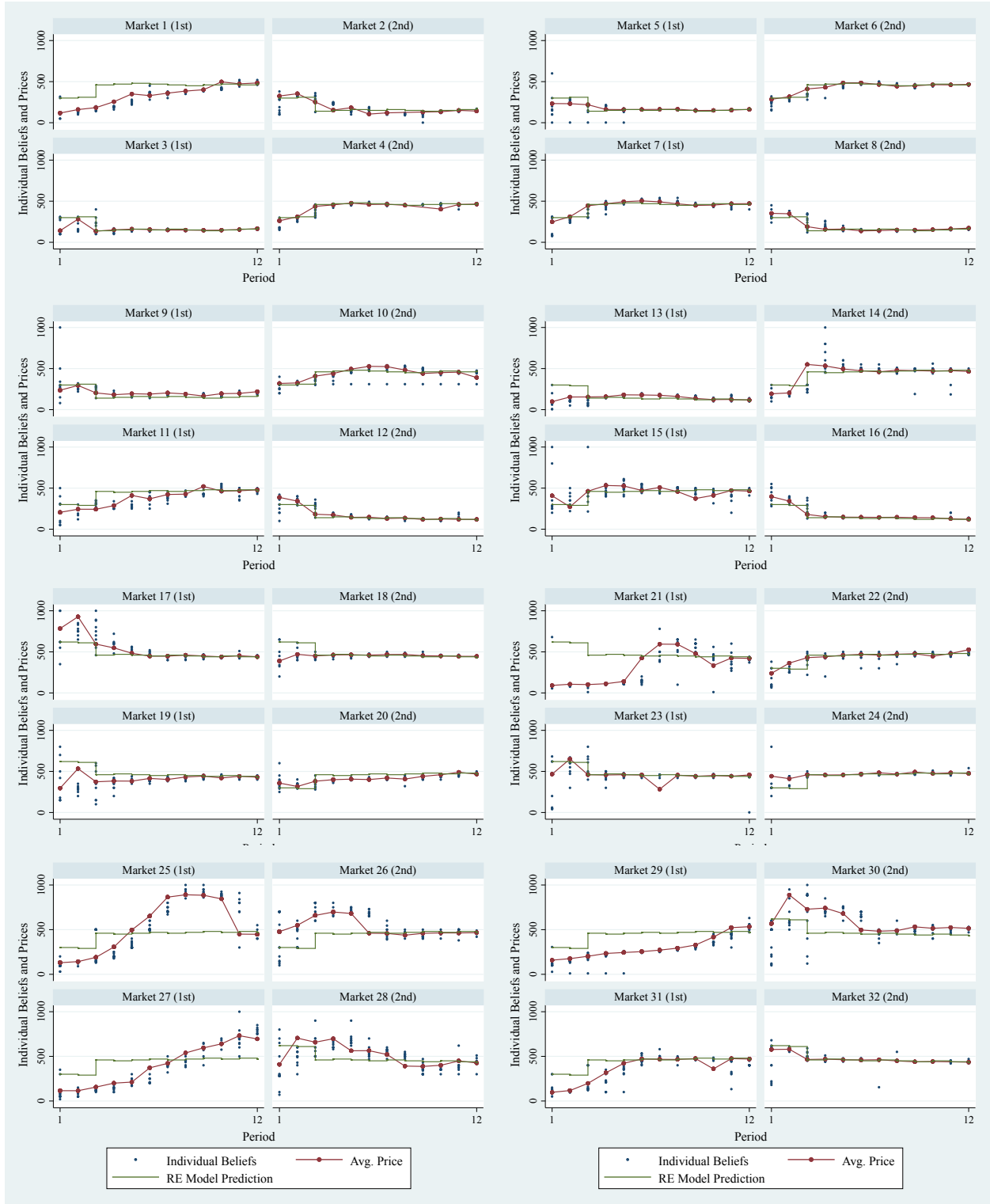


Fig. 3. Overview of market prices and individual beliefs in all markets
 This figure depicts average transaction prices and individual beliefs relative to Rational Expectations levels in each period of each session. The individual graph titles also indicate whether a market was first or second in a session.

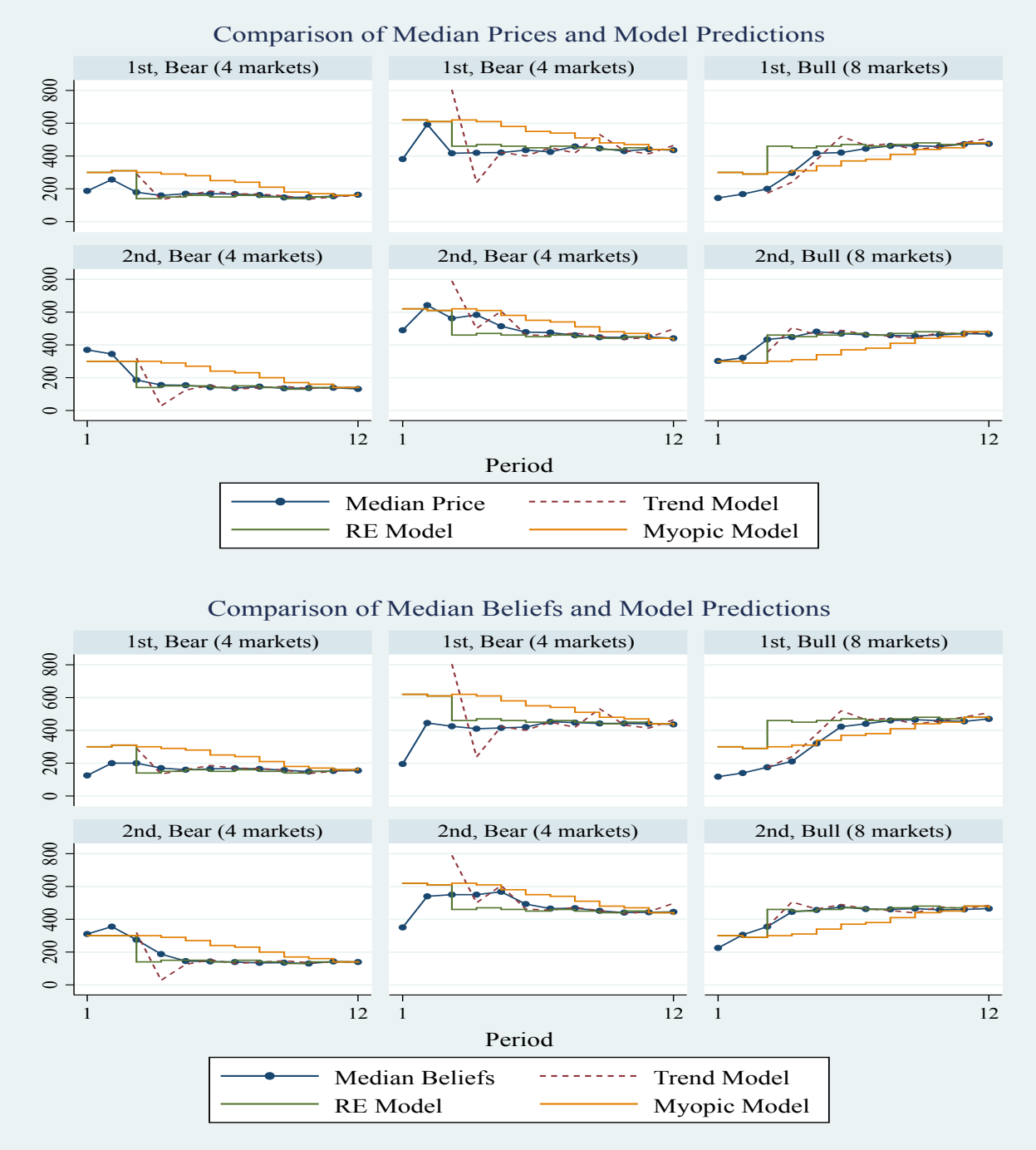


Fig. 4. Comparison of market prices and beliefs against model predictions
 The upper panel plots median market prices, separately for first and second, as well as Bull and Bear markets. Additionally, predictions for our different models are shown: the Trend Model, the RE Model and the Myopic Model. The lower panel plots the medians of subjects beliefs against these model predictions.

the RE Model achieves the best fit ($\text{RMSE} = 52.25$) compared to the Trend Model ($\text{RMSE} = 95.84$) and the difference is both economically and statistically significant. Comparing model fit in first and second markets implies that subjects also seem to learn about the fundamental value process relatively quickly and this increasingly reflected in market prices. This finding is particularly noteworthy since we do not keep parameters unchanged from the first market to the second and thereby make the conditions for learning more difficult. Apparently, prices seem to be more likely to follow trends in Bull than in Bear markets when traders are inexperienced. Finally, the Myopic model performs relatively poorly in describing prices. These patterns constitute the basis of our first result:

Result 1: The Rational Expectations Model describes prices more accurately than the Trend or the Myopic models.

Next, we turn to a comparison of the models with regard to how well they predict beliefs. The data are shown in Panel B of Table 2. Essentially, the results we obtained from the models' fit to prices also translate to beliefs. The RE model performs better in Bear markets ($\text{RMSE} = 78.07$), while the Trend model is more accurate in Bull markets ($\text{RMSE} = 77.44$). The Myopic Model again performs poorly relative to the other two. All three models are more accurate in market 2 than in market 1, reflecting the greater structure and lower noise in the belief data in the second market. The RE Model achieves the best fit in the second market ($\text{RMSE} = 70.32$) followed by the Trend Model ($\text{RMSE} = 77.70$). The difference between the RMSE of both models is statistically significant but less pronounced when considering beliefs rather than prices. This serves as a first hint that there might be heterogeneity in the way subjects form beliefs. The main pattern is summarized as our second result.

Result 2: The Rational Expectations and Trend Models describe beliefs more accurately than the Myopic Model.

Table 2: RMSE of different model predictions

Panel A: Prices						
	I. RE	II. Myopic	III. Trend	Diff I-II	Diff I-III	Diff II-III
Bull	86.62	109.97	79.52	-23.34 (0.148)	7.09 (0.956)	30.44** (0.034)
Bear	52.25	96.9	95.84	-43.59*** (0.006)	-44.65*** (0.008)	1.06 (0.605)
1st Market	94.56	114.24	98.71	-19.68 (0.214)	-4.15 (0.877)	15.53 (0.400)
2nd Market	44.31	92.64	76.67	-48.32*** (0.006)	-32.34*** (0.002)	15.97 (0.148)
Panel B: Beliefs						
	I. RE	II. Myopic	III. Trend	Diff I-II	Diff I-III	Diff II-III
Bull	112.31	116.12	77.44	-3.81 (0.796)	34.86* (0.098)	38.68*** (0.004)
Bear	78.07	108.02	98.74	-29.95** (0.015)	-20.67*** (0.007)	9.27 (0.470)
1st Market	120.06	127.06	98.49	-7.25 (0.756)	21.57 (0.534)	28.82 (0.098)
2nd Market	70.32	96.83	77.7	-26.51** (0.023)	-7.38* (0.063)	19.13** (0.044)

This table presents the observed deviations of prices (Panel A) and beliefs (Panel B) from the predictions implied by the following models: Rational Expectations Model (RE), Myopic Model (Myopic) and Trend Model (Trend). In the three columns on the left, we calculate the root mean squared errors (RMSE) at the individual market level (16 markets per cell) and present the averages over different treatments. In the three columns on the right, we provide the results from Wilcoxon signed-rank tests on the Null that differences between RMSEs are zero: The p-values are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.3. Aggregate market behavior

We have shown that the Rational Expectations Model describes the price paths reasonably well, particularly after repetition. We now consider the correlates of the deviations from the RE benchmark that we observe. We estimate the following regression of aggregate market-level mispricing, measured in RAD, on a number of potential correlates. These are dummy variables for treatment and first market of a session, as well as the mean and standard deviation of CRT scores in the group of traders participating in the session. Each market j corresponds to one observation.

$$\text{RAD}_j = \beta_0 + \beta_1 \times \text{Bull}_j + \beta_2 \times \text{FirstMarket}_j + \beta_3 \times \text{Avg.CRT}_j + \beta_4 \times \text{SD.CRT}_j + \epsilon_j \quad (8)$$

We also estimate a similar specification for turnover.

$$\text{Turnover}_j = \beta_0 + \beta_1 \times \text{Bull}_j + \beta_2 \times \text{FirstMarket}_j + \beta_3 \times \text{Avg.CRT}_j + \beta_4 \times \text{SD.CRT}_j + \epsilon_j \quad (9)$$

Table 3 shows aggregate efficiency statistics across treatments.¹⁰ The regression results for equations 8 and 9 are shown in Table 4. *Mispricing* refers to deviations of prices from the RE model predictions, since the RE is a model assuming full rationality on the part of traders. Table 3 indicates the value of the measures of mispricing and allows Hypotheses 1 and 2 to be evaluated.

In addition, since both hypotheses conjectured similar mispricing across first and second markets, as well as across Bull and Bear markets, we conduct *tests of equivalence*.¹¹ Under equivalence tests, the role of the Null Hypothesis and the burden of proof are reversed, because the Null now conjectures that sample statistics are not equivalent. Equivalence is established if the average RADs of two samples are close enough to one another such that one cannot be considered inferior or superior to another (Walker and Nowacki, 2011). Formally, equivalence is established at the 5% significance level if the 90% confidence interval for the difference in RAD lies entirely within the interval $(-\delta, \delta)$, where δ represents the *equivalence margin*. We set δ equal to 10%. The difference in RAD between first and second markets is -10.89% and the 90%-confidence interval is [-17.60%, -4.18%], which is partially outside the interval (-10%, 10%). Thus, we reject Hypothesis 1 that prices are equally close to RE levels in first in second markets. The difference in RAD between Bull and Bear markets is only 0.18% and the 90%-confidence interval is [-7.32%, 7.7%]. Therefore, we cannot reject

¹⁰Table 12 in Appendix B provides more detailed data for each individual market.

¹¹We thank a referee for drawing our attention to such tests, which are commonly used in clinical research.

Hypothesis 2 at the 5%-level, so that Bull and Bear markets are deemed equally efficient.¹²

Table 4, in the first three columns, reports regressions of price efficiency on a number of potential determinants. The table shows that the coefficient of the variable *FirstMarket*, a dummy variable equal to 1 if the market was the first conducted in a session, is positive and significant in each of the first three columns. In contrast, the dummy variable for treatment *Bull*, is not significant under any specification. Thus, we state our next two results, which report our evaluation of hypotheses 1 and 2.

Table 3: Aggregate efficiency measures

	RAD	RD	GD	GAD	Turnover	Vola
Mean Bull	16.55%	-3.69%	-7.43%	23.91%	0.25	15.89%
Mean Bear	16.74%	1.20%	0.00%	20.55%	0.25	12.91%
Mean 1st	22.09%	-7.22%	-11.12%	33.07%	0.31	19.05%
Mean 2nd	11.20%	4.73%	3.69%	11.39%	0.18	9.76%
Median Bull	9.70%	-0.10%	-2.63%	11.47%	0.20	10.04%
Median Bear	13.80%	0.10%	0.04%	12.29%	0.24	10.41%
Median 1st	16.72%	-8.26%	-6.43%	23.02%	0.29	18.03%
Median 2nd	8.93%	1.63%	-0.22%	9.42%	0.16	8.89%

This table provides aggregate mispricing measures across treatments. The mispricing measures are calculated as follows: Relative Absolute Deviation (RAD) = $\frac{1}{T} \sum_{t=1}^T \frac{|\bar{P}_t - RE_t|}{\overline{RE}_m}$, where \bar{P}_t is the average price in a period, RE_t is the rational expectations value in that period, and \overline{RE}_m is the average rational expectations level in a market. RD is the Relative Deviation, defined as $\frac{1}{T} \sum_{t=1}^T \frac{\bar{P}_t - RE_t}{\overline{RE}_m}$. Geometric Absolute Deviation (GAD, Powell (2016)) = $\exp(\frac{1}{T} \sum_{t=1}^T |\ln(\frac{\bar{P}_t}{RE_t})|)$. GD is the analogous geometric deviation. Turnover is defined as the total trading volume per period, divided by the total number of shares held by all traders. Vola measures the price volatility within periods and is defined as: $Vola = \frac{1}{T} \sum_{t=1}^T \sqrt{\frac{1}{N} \sum_{n=1}^N (RET_n - \overline{RET}_t)^2}$ where $RET_n = \ln(\frac{P_n}{P_{n-1}})$.

Result 3: Prices are closer to RE levels in Market 2 than in Market 1.

Result 4: Mispricing does not differ between Bull and Bear markets.

We also consider differences in Turnover, our measure of transaction volume, between the first and second markets. In the Smith et al. (1988) paradigm, turnover declines as the same traders repeat the market. Decreasing volume can be taken as a sign of less disagreement

¹²Non parametric signed-rank tests of differences in RAD between the first and second markets reveals that mispricing is lower in the second market than in the first, but does not differ between Bull and Bear treatments.

Table 4: Determinants of Mispricing and Turnover

	(1)	(2)	(3)	(4)	(5)	(6)
	RAD	RAD	RAD	Turnover	Turnover	Turnover
Bull	-0.0019 (-0.05)	-0.0020 (-0.05)	-0.0053 (-0.12)	-0.0009 (-0.02)	0.0135 (0.28)	0.0071 (0.15)
First Market	0.1089** (2.71)	0.1141** (2.57)	0.1137** (2.54)	0.1246** (2.72)	0.1333** (2.80)	0.1287** (2.65)
Avg. CRT		-0.0054 (-0.14)			-0.0503 (-1.40)	
SD. CRT			0.0615 (0.89)			0.0193 (0.29)
Constant	0.1130** (3.94)	0.1383 (0.78)	-0.0034 (-0.02)	0.1870** (5.63)	0.4257** (2.60)	0.1624 (1.20)
R^2	0.2020	0.2065	0.2324	0.2037	0.2784	0.2290
Observations	32	28	28	32	28	28

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$

This table presents the results from a regression of RAD (columns 1 - 3) and Turnover (columns 4 - 6) on a number of covariates. Each market is an observation. Bull is a dummy variable that equals 1 for Bull markets. First Market is a dummy that equals 1 if the market is the first in the session. Avg. CRT represents the average cognitive reflection test score of a cohort in a given market. SD. CRT represents the standard deviation of scores within a cohort and captures the heterogeneity in CRT scores within a market. We use Huber-White robust standard errors. Columns (1) and (4) include observations from session 1 and 2, where we measured CRT on a 3-item scale.

among traders regarding appropriate prices, and thus as an indication that the market has stabilized. We observe that Turnover decreases between markets 1 and 2, as the prices move closer to RE levels. This can be seen in columns 4 - 6 of Table 4.

There are also consistent patterns in the time profile of Turnover within the life of the asset. Turnover decreases towards the end of a market. This is consistent with subjects' convergence towards homogeneous beliefs, and a lower need for risk-sharing due to the decreasing uncertainty about final earnings as the end of the market approaches. As shown in Table 12 in Appendix B. Turnover is highest in the first period and in the third period, right after the announcement of prevailing business conditions. This Turnover pattern underscores the significance of our design choice to include two neutral periods before the onset of a trend. It gives subjects the chance to change their exposure to the risky asset during the first two periods, knowing that a shock to the earnings process will occur. Turnover is not significantly higher in Bear than in Bull markets. The patterns of turnover are summarized as our fifth result.

Result 5: Turnover decreases over time, both within and between markets, and peaks after the announcement of the trend.

Since our price trajectories are relatively well described by Rational Expectations, the question arises of how this price discovery is achieved. A candidate explanation suggested by the previous literature might be that relatively sophisticated subjects drive prices closer to RE levels. Thus, we would expect markets with a higher cohort average CRT score to exhibit prices closer to RE levels. The previous literature, however, is somewhat inconclusive about the relationship between CRT and mispricing. Some studies show that the presence of a greater share of sophisticated subjects correlates with less mispricing (Corgnet et al., 2015; Bosch-Rosa et al., 2018; Breaban and Noussair, 2015; Charness and Neugebauer, 2018; Cueva and Rustichini, 2015). Hanaki et al. (2017), however, construct trader cohorts for whom CRT scores are homogeneous and heterogeneous, and find that efficiency is lowest for heterogeneous groups. They argue that sophisticated subjects are aware of the presence of noise traders and try to manipulate the market. Thus, we test for significance of both the average CRT-level, as well as the standard deviation of CRT, to account for such non-monotonicities.

The regression in Table 4 reveals that markets with higher cohort CRT scores do not price significantly closer to RE levels. Prices and quantities traded also exhibit no relationship to the standard deviation of CRT scores. Prices in our design adhere closely to rational expectations even when CRT scores are lower, indicating that this pattern is robust to the

inclusion of relatively unsophisticated subjects. This holds at least for the range of CRT scores that we observe in our sample.

Result 6: Adherence of prices to rational expectations is uncorrelated with group average CRT score.

4.4. Evidence of under- and overreaction

Thus far, we have described the general properties of our markets, such that our measures of aggregate efficiency can be compared with existing experimental designs that study price discovery in long-term asset markets. Now, we turn to the question of whether price dynamics in our setting are consistent with behavioral models of under- or overreaction. The three models considered applicable to our setting predict different price patterns: The BSV model predicts initial underreaction to news shocks and subsequent overreaction, whereas the model by Frazzini (2006) predicts initial underreaction and subsequent convergence. Conversely, the efficient market hypothesis, represented by Fama (1998)'s critique of such behavioral models, predicts under- and overreaction to be equiprobable and non-systematic. By eyeballing individual price trajectories in Figure 3 again, it is possible to identify markets that are rather consistent with the rationale of the BSV model (e.g. markets 2, 10, 25), as well as markets that rather consistent with Frazzini (2006)'s rationale (e.g. markets 9, 11, 20). However, to abandon Rational Expectations in favor of any of the two behavioral alternatives, one model would have to consistently outperform overall. Otherwise, our markets would be regarded as efficient *on average*. We next study whether either of the two models pass this test.

Table 5 displays the single-period relative deviation of prices from fundamentals (RD) before and after the news shock at the end of period 2, separately for Bull and Bear markets. Each value represents the average RD in a particular period from pooling 16 markets in each treatment. We also test whether the average RD in a period is significantly different from zero. If price dynamics in Bull markets conformed to the BSV model, we should observe statistically significant underpricing ($RD < 0$) shortly after period two and overpricing ($RD > 0$) in later periods. For Bear markets, the BSV model predicts opposite signs. If, however, the model by Frazzini (2006) better described price patterns, we should observe consistent overpricing in Bear markets and consistent underpricing in Bull markets, at least in the periods just after the news shock.

Table 5 shows that the asset is underpriced in the first two periods before the news shock, which might be due to risk-averse subjects disposing of their assets. However, by the second

period, asset prices have converged to fundamental values and are no longer significantly different from zero. This finding underscores the importance of our decision to include two neutral periods in the beginning, because it allows us to study news reactions when the market is close to equilibrium. It also suggests that two periods was long enough for traders to accomplish their desired initial portfolio rebalancing. Regarding news reactions, we start with Bull markets and begin by looking at the signs of RD first. Table 5 seems to support the BSV model: Prices on average significantly underreact to the positive news in periods 3 and 4, monotonically approach fundamentals from below until period 6, and tend to overshoot fundamentals thereafter. This appears in line with our earlier observation that the Trend Model provides a fairly accurate description of price dynamics in Bull markets that consist of inexperienced traders. However, the values lose statistical significance after period 4 and are also remarkably close to zero in magnitude. In fact, our results for Bull markets are reminiscent of the short-term event-study literature that typically finds (non-exploitable) underreaction in the first days after the announcement and random-walk-like behavior thereafter (Fama et al., 1969; Bernard and Thomas, 1989). This pattern also seems to apply to our data. Thus, our results for Bull markets do not tip the balance against the hypothesis that the market prices positive news efficiently on average. Our results regarding the market response to positive news are stated as Result 7.

Result 7: Positive news is priced efficiently on average.

Regarding Bear markets, the evidence is more nuanced. Firstly, since the average RD never flips signs, the BSV-model receives no support in Bear markets. Secondly, there are more periods in which the asset is on average significantly overpriced. Such a pattern is more consistent with Frazzini (2006)'s model, which predicts that traders who suffer from the disposition effect are reluctant to realize losses after negative news shocks. However, two observations speak against an argument linking the disposition effect to price behavior. The first is that price patterns in Bear markets are not as monotonic as in Bull markets, indicating that the form that mispricing takes in the Bull market has more structure. The second is that trading volume patterns, analyzed in Table 4, do not support an explanation based on the disposition effect, which requires higher trading volume in Bull than in Bear markets. This is because traders ought to realize gains prematurely after positive news and exhibit reluctance to realize losses after negative news. Result 8 concerns the reaction of the market to negative news and draws a similar conclusion to Result 7.

Result 8: Negative news is priced efficiently on average.

Table 5: Underreaction patterns in Bull and Bear markets

Period	Bear		Bull	
	RD	p-values	RD	p-values
1	-26.2%*	0.098	-10.6%	0.148
2	-3.1%	0.877	-5.8%	0.365
3	16.3%**	0.049	-22.0%**	0.015
4	4.7%	0.133	-12.1%*	0.079
5	3.5%*	0.066	-5.0%	0.605
6	4.7%	0.134	-5.0%	0.379
7	0.1%	0.958	2.6%	0.776
8	6.1%**	0.047	2.1%	0.552
9	4.1%**	0.044	2.4%	0.433
10	1.0%	0.339	2.8%	0.205
11	2.5%	0.320	3.5%	0.938
12	3.0%	0.127	3.5%	0.717

This table presents the average Relative Deviation (RD) over periods across Bull and Bear markets. Each value is based on 16 markets. Positive RD indicates overpricing and negative RD indicates underpricing. The p-values refer to Wilcoxon signed-rank tests of the Null that RD over all 16 markets in that period equals zero (* $p < 0.10$, ** $p < 0.05$).

Most importantly, following Fama’s line of reasoning, it would only be admissible to reject the Efficient Market Hypothesis in favor of an alternative, if that alternative were to outperform EMH in *both* Bull and Bear markets and not only under specific instances. Neither of the two behavioral alternatives passes this test in our data. Finally, prices are on average so close to fundamental values that we cannot reject the part of hypothesis 3 that asserts that Relative Deviation is equal to 0. We will re-evaluate the extent of mispricing more rigorously in Section 5, where we embed our results in a formal meta-study framework and compare our level of mispricing with that of related asset market designs.

4.5. *Determinants of belief formation and accuracy*

We now turn to an analysis of beliefs. We first analyze how the accuracy of subjects’ price predictions varies with their cognitive sophistication, as measured by their CRT-score. It seems plausible that subjects with higher cognitive ability would have a superior understanding of the fundamental value process. Thus, to the extent that market prices accurately reflect fundamental values, more sophisticated subjects should predict prices better. However, in markets where prices temporarily decouple from fundamental values a capability to “read others’ mind” and to anticipate price movements might be required. Within the

SSW-design, Bosch-Rosa et al. (2018) studied the variation in prediction accuracy across distinct dimensions of sophistication that were elicited with different tests: cognitive reflection, strategic sophistication and backward-induction capabilities. They found all three dimensions to be important, particularly in less efficient markets. Our experimental design can add another perspective because it allows us to clearly distinguish between rational, trend-following and myopic expectations. We can therefore test how subjects with different degrees of cognitive sophistication adhere to different models of expectation formation.¹³

Table 6 regresses subjects’ average relative prediction errors ($\frac{1}{12} \sum_{t=1}^{12} \frac{|b_t - \tilde{P}_t|}{\tilde{P}_t}$) on their CRT scores, their study backgrounds and on treatment dummies. The regression reveals that subjects’ predictions improve from the first to the second market. Prediction accuracy is also increasing in cognitive ability, particularly so in second markets. In terms of economic magnitude, however, sophisticated subjects are only marginally better at price prediction than their less sophisticated counterparts. This might be an indication of the overall efficiency of our markets, as well as of the fact that cognitive ability alone is insufficient to anticipate price movements. Interestingly, prediction accuracy does not vary between Bull and Bear markets.

Result 9: Prediction accuracy increases from first to second markets. Sophisticated subjects predict prices marginally better.

To shed further light on this issue we next analyze how the adherence to different models of expectation formation varies with cognitive ability. Table 7 regresses subjects’ RMSEs with regard to our three models on their CRT scores, their study background and on treatment dummies. The regression reveals an interesting pattern. In first markets, subjects with higher CRT scores adhere more closely to the RE Model and less closely to the Trend Model. However, in second markets sophisticated subjects seem to adapt: while they on average still adhere closer to the RE Model than less sophisticated subjects, they now also adhere closer to the Trend Model. Thus, sophisticated subjects seem to have learned about the market environment, and in particular seem to have learned that relying on the Rational Expectations model alone is suboptimal for predicting prices. Instead, they now also take contemporaneous price trends into account. Finally, cognitive ability does not have any discriminatory power with respect to the Myopic Model. This is because the Myopic Model

¹³Other experiments requiring backward induction, such as Carpenter et al. (2013), suggest that we can expect more sophisticated subjects to form their beliefs more in accordance in Rational Expectations. However, an accurate prediction of price dynamics requires more than a superior understanding of the underlying economic theory. In particular, recent studies have shown that in addition to numerical or cognitive capabilities, mentalizing (“perspective-taking”) capabilities might be necessary to master the complex dynamics in asset markets (Bruguier et al., 2010; Hefti et al., 2016).

Table 6: Prediction accuracy and CRT-scores

	1st Markets Prediction Error	2nd Markets Prediction Error
CRT	-0.00654** (-2.35)	-0.00978*** (-3.84)
Econ	-0.00122 (-0.07)	-0.00546 (-0.49)
Bull	0.00594 (0.21)	-0.0204 (-0.93)
Constant	0.174*** (6.06)	0.126*** (8.57)
Observations	126	126
R^2	0.023	0.166

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In this table we regress subjects' average relative prediction error on a number of covariates. Each participant is the unit of observation. CRT represents a subject's CRT score and is centered at the mean. Econ is a dummy variable that equals one for subjects with an Economics or Business Administration background. Bull is a dummy variable that equals one if the market was from the Bull treatment. Standard errors are clustered at the market level.

Table 7: Model fit to beliefs as a function of CRT-scores and other variables

	RMSE of Beliefs in 1st Markets			RMSE of Beliefs in 2nd Markets		
	RE	Trend	MM	RE	Trend	MM
CRT	-5.548** (-2.43)	3.303* (2.04)	-0.360 (-0.14)	-5.941** (-2.89)	-4.210* (-1.82)	0.938 (0.53)
Econ	19.74 (1.42)	11.51 (1.52)	17.90 (1.12)	3.329 (0.35)	9.525 (1.31)	19.06 (1.49)
Bull	65.61 (1.56)	-15.13 (-0.59)	2.475 (0.06)	1.510 (0.06)	-26.87 (-1.13)	19.02 (0.84)
Constant	74.09** (2.45)	98.28*** (4.60)	119.5*** (3.27)	65.33*** (3.44)	82.45*** (4.28)	77.23*** (9.71)
Observations	126	126	126	126	126	126
R^2	0.167	0.0484	0.0118	0.0485	0.0851	0.0874

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In this table we regress subjects' RMSE for our three models of belief formation on a number of covariates, separately for first and second markets. CRT represents a subject's CRT score and is centered at the mean. Econ is a dummy variable that equals one for subjects with an Economics or Business Administration background. Bull is a dummy variable that equals one if the market was from the Bull treatment. Standard errors are clustered at the market level.

has the poorest fit among the three models, and neither sophisticated nor unsophisticated subjects seem to form their beliefs in accordance with it.

Result 10: In first markets, sophisticated subjects adhere more closely to the RE model and less closely to the Trend and the Myopic models than less sophisticated traders.

5. Discussion

The majority of experiments studying price discovery in markets for long-lived assets has in some form or another relied on the pioneering design by Smith et al. (1988). The SSW-design has thereby become a standard paradigm in experimental finance. Yet, for various reasons, the SSW-design has not remained unchallenged. Two major points of criticism are its alleged lack of realism (Oechssler, 2010) and the confusion the declining fundamental value trajectory appears to cause among subjects (Kirchler et al., 2012). Such caveats raise the question as to what extent experimental findings - such as a frequent decoupling of prices from fundamental values - generalize to capital markets in the field.

Our design, like the others in the experimental literature, is highly stylized compared to those in the field, and thus we do not fully address the issue of realism. Our setup does include an important feature that the SSW structure leaves out, which is the arrival of news affecting the fundamental value. This allows the market the freedom to under-or over-react to the arrival of news. While we are not the first to study an experimental market with this feature, we feel that it is an important one to include. Our experimental design is adapted from a setting in the finance literature that has long been exploited to study price discovery: analyzing share price reactions to corporate announcements (Ball and Brown, 1968; Fama et al., 1969). Our design also allows us to clearly distinguish between competing models of expectation formation. Our results are at odds with conventional wisdom from previous asset market studies: Prices in our experiment adhere remarkably closely to Rational Expectations levels. We do not observe a widespread tendency for bubbles to form.

The good performance of the Rational Expectations model suggests that there is less confusion in our experiment than in some others. We believe that the stationary phase at the beginning of the market makes it more likely that the environment would be well understood by participants. It is known from prior research (e.g, Kirchler et al., 2012) that environments with stationary fundamental values lead to less confusion. Furthermore, the interaction of traders with different levels of sophistication early on in the market, with more sophisticated individuals presumably more likely to be on the profitable side of trades, means that at the time of the onset of the trend, relatively sophisticated traders will hold

on average more than their share of the endowment and cash. These traders are then in a better position to influence the market subsequently, when the trend begins. Furthermore, the news shock in our experiment is salient and affects the level of all future earnings. This may encourage traders to form forward-looking expectations.

Replication of established experimental results is increasingly regarded as a necessary exercise, as a recent collaborative effort of major experimental economists has demonstrated (Camerer et al., 2016). We make no claims here that any previous experiment does not replicate when the same experiment is conducted again. However, replication does not only concern running a previous experiment under identical conditions, but also checking whether prior results are obtained under related though different settings (Maniadis et al., 2014) to verify that a previously obtained result is robust. In experimental asset market research, like empirical asset pricing research (Fama, 1970), there is a *joint hypothesis problem*, or as Smith (2002) puts it: “experimental results always present a joint test of the theory that motivated the test, and all the things you had to do to implement the test”. Consequently, if we seek to use the experimental method to address questions regarding the informational efficiency of capital markets, it is risky to base our conclusions on such a limited set of market designs. Having created a research design which supports the Rational Expectations Hypothesis, the question arises whether and how our beliefs about efficiency in experimental asset markets ought to be revised.

To shed light on this issue, we collected data from a variety of previous asset market experiments with different designs and compared the results with respect to the relative deviation (RD) of average market prices from fundamental values. While RD is a somewhat crude measure to capture price dynamics in markets, it is relatively straightforward to extract from existing studies and provides a reasonable measure of the extent of under- or overpricing across different market designs. For a meaningful comparison, we included a recent replication of the SSW-design (Kirchler et al., 2012), a design with both deterministically increasing and decreasing fundamentals (Stöckl et al., 2014), and a design with delayed onset of the increasing or decreasing trajectory (Breaban and Noussair, 2015). We also included the results by Lin and Rassenti (2012), who also employ stochastic fundamentals, and whose design has the closest similarity to ours.

Figure 5 presents the average relative deviation of market prices from fundamentals and the estimated 95%-confidence intervals across different designs. Two prominent patterns emerge. Firstly, previous experimental evidence suggests that assets tend to be overvalued in markets with decreasing fundamentals and tend to be undervalued in markets with increasing fundamentals. Secondly, wide confidence intervals imply that most markets in previous designs were heterogeneous in efficiency. Moreover, the confidence intervals of RAD do not

contain zero. This implies that most previous studies rejected the null hypothesis that market prices on average accurately reflected fundamental values. In contrast, the markets in our study were relatively homogeneous, since we rarely observed a severe decoupling of prices from fundamentals. More importantly, we cannot reject the null hypothesis that markets are on average efficient in our design.

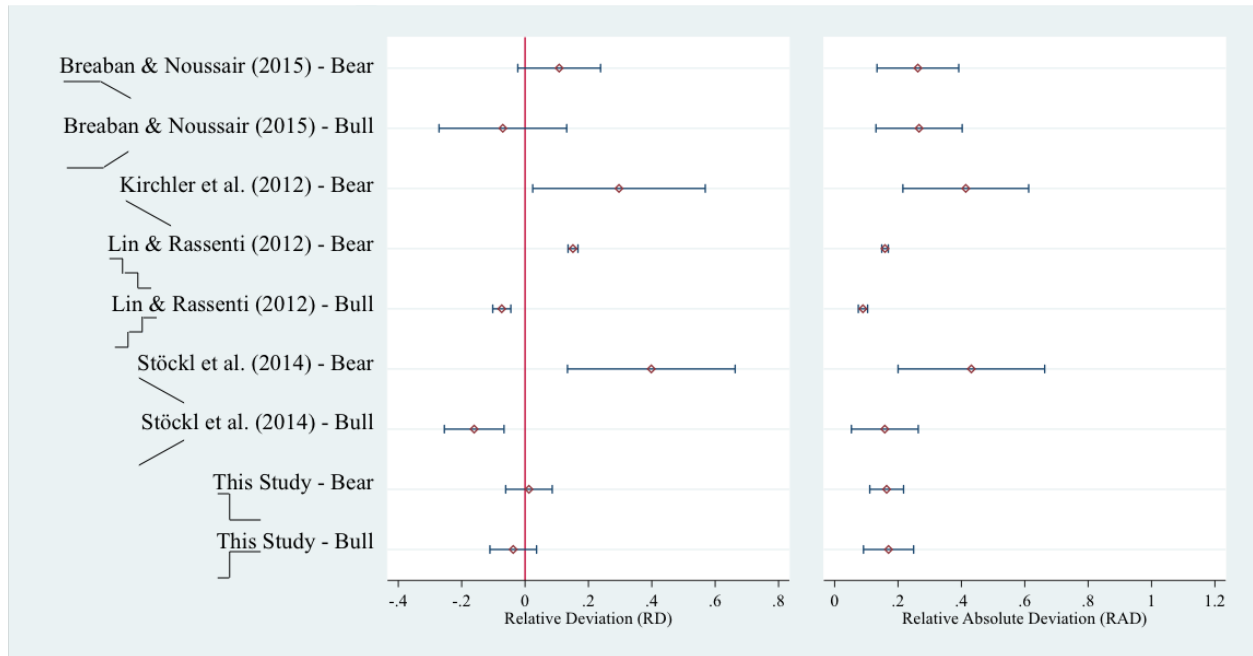


Fig. 5. Over- and underpricing across different asset market designs

This figure compares over- and underpricing across different asset market designs. For each study, we compute the Relative Deviation (RD) and Relative Absolute Deviation (RAD) of prices from fundamental values at the individual market level. We compute the average RD and RAD over all markets with a particular design and provide 95%-confidence intervals. Values of RD above 0 indicate overpricing and values below 0 indicate underpricing. Values of RAD above 0 indicate mispricing. Number of observations (i.e. markets) in each study from top to bottom: 20, 12, 6, 8, 8, 6, 6, 16, 16. The original SSW-experiment is replicated in the study by Kirchler et al. (2012).

To reconcile our results with previous studies, we next evaluate our results in a meta-scientific Bayesian framework. Ioannidis (2005) and Maniadis et al. (2014), among others, urge researchers not only to consider the p -values of their studies' results in isolation, but also to put their results more into perspective. In particular, whenever researchers discover new scientific relationships, replicate or fail to replicate established findings, they ought to incorporate their *research prior* (how likely a certain finding would be considered ex-ante) and the *statistical power of their design*. Integrating these elements yields a finding's *post-study probability* (PSP), i.e. the posterior probability of an experimental effect being true.

The PSP after a successful replication of a previous experimental result is defined as follows:

$$PSP^+ = \frac{(1 - \beta)\pi}{(1 - \beta)\pi + \alpha(1 - \pi)} \quad (10)$$

where $(1 - \beta)$ refers to the statistical power, π refers to the research prior about a proposed relationship being true, and α is the typical significance level in the field. Thus, PSP^+ is determined by the probability of correctly rejecting the null hypothesis (that an alleged effect does not exist), the prior belief that an effect exists, and the power of the test used to identify this effect. Conversely, the post-study probability after failing to replicate a certain effect, or PSP^- , is defined as:

$$PSP^- = \frac{\pi\beta}{\pi\beta + (1 - \pi)(1 - \alpha)} \quad (11)$$

Prior experimental evidence stresses the relative ease with which prices may decouple from fundamental values in long-term asset markets. We are interested in how this prior belief π ought to change in light of our results. The following calculation provides a first qualitative answer to this question and is informative about approximate magnitudes.

The relative overpricing in previous studies of markets with declining fundamental value time trajectories, which we will call Bear markets, was on average 21.67% of fundamental value (based on 40 markets), which represents a statistically significant deviation from $H_0 : RD = 0$, or efficient pricing, at the 1%-level. Bull markets were on average only -8.5% underpriced (based on 26 markets), and this difference is only significantly different from 0 at the 10%-level. In our study, we fail to replicate such levels of over- and underpricing and cannot reject the null hypothesis of efficient pricing. Running 16 markets for each treatment, our design achieves 80% power for identifying such overpricing in Bear markets, but only 29% power for identifying underpricing in Bull markets.¹⁴ Table 8 demonstrates how the lack of pronounced mispricing in our design yields different research posteriors (PSP^-) for different strengths in initial priors (π). Table 8 suggests that belief revision should be strongest for our priors about market efficiency in bear market designs, due to our relatively high statistical power in that setting. For instance, if one held a strong prior belief, of say

¹⁴Our post-hoc power analysis assumes that the estimates from previous experimental evidence for mean and standard deviation of RD are equal to the true population mean and standard deviation. We weight each study equally. This strikes us as reasonable since the number of markets conducted in the prior studies was comparable: Twelve and twenty sessions in Breaban and Noussair (2015) Bull and Bear treatments, eight each in Lin and Rassenti (2012) Bear and Bull treatments, six each in the Bull and Bear treatments of Stöckl et al. (2014), and six in the treatment of Kirchler et al. (2012) that we used in the analysis. There are 16 markets in each of the two treatments in our study.

80%, that experimental asset markets with declining fundamentals generate overpricing of about 21.67% relative to fundamental values, this belief should be revised downwards to 48% given our findings. Our evidence has less drastic effects for bull market designs given the inconclusiveness about the extent of mispricing in such markets from previous studies and our relatively low statistical power.

Table 8: PSP estimates as a function of prior probability

Bear		Bull	
Prior π	PSP^-	Prior π	PSP^-
0.1	0.03	0.1	0.07
0.2	0.06	0.2	0.15
0.3	0.09	0.3	0.23
0.4	0.14	0.4	0.31
0.5	0.19	0.5	0.41
0.6	0.26	0.6	0.51
0.7	0.35	0.7	0.61
0.8	0.48	0.8	0.73
0.9	0.68	0.9	0.86

6. Modeling convergence towards rational expectations: adaptive learning

In the previous sections we have established that prices and beliefs in our experimental setting adhere more closely to Rational Expectations than in comparable experimental settings. Nevertheless, in some of our markets, we do observe patterns that are inconsistent with Rational Expectations. In particular, subjects seem to form expectations differently in those markets in which prices slowly gravitate towards or overshoot Rational Expectations levels before eventually converging. Experimental research ideally informs theoretical modeling and vice versa. Therefore, we advance an *Adaptive Learning Model* to capture the convergence of subjects' beliefs towards Rational Expectations.

Models in this tradition acknowledge that Rational Expectations are an equilibrium outcome and try formalize the process whereby actors recursively learn the equilibrium parameters. The umbrella term “adaptive learning” covers a variety of models, such as *reinforcement learning*, *belief based learning*, and *genetic algorithms*.¹⁵ Generally speaking, these models assume that cognitive constraints prevent agents from performing the necessary complicated calculations to arrive at equilibrium parameters. Instead, agents have a set of models, actionable strategies or heuristics in mind and choose among these depending on the models' past relative performance. Adaptive learning models have been successfully estimated on standard experimental games (e.g. Erev and Roth 1998) and have shed light on asset pricing anomalies, which are difficult to reconcile with standard rational expectations (e.g. Hong et al. 2007; Adam et al. 2016).

Subjects' beliefs in our experiment might deviate from rational expectations for at least two reasons. On the one hand, they might be boundedly rational and misperceive the fundamental value process underlying the asset in our experiment. Over time, these subjects learn about the appropriate model from observed prices. On the other hand, despite having the correct underlying model in mind, they might question the rationality of other traders. This induces higher order uncertainty and might make it optimal to state price expectations that deviate from fundamental values. For instance, Corgnet et al. (2018) show that trading performance in SSW experiments is not only determined by the knowledge of economic theory but also by the ability to anticipate other traders' motives. Under either scenario, the fundamental news shock in our experiment might not be incorporated instantaneously into prices.

The intuition of our Adaptive Learning Model, hereafter ALM, is as follows. For sim-

¹⁵Duffy (2006) provides a comprehensive overview of these learning models and the attempts to estimate them on experimental data. See Camerer (2011) for behavioral and psychological foundations of these models.

plicity, we assume that subjects have three distinct models of price formation already in their mind: Rational (RE), Myopic (ME) and Trend-following expectations (TR). Learning follows an iterative process, in which subjects assign changing weights to these models according to the models' past predictive performance. Models that have performed well in the past receive relatively higher weights compared to models that have performed badly. One might also conceive of additional models, but the use of a larger number of potential models exceed subjects' cognitive capacities in keeping track of the models' performance. To form a prediction, subjects compute the weighted average of different models' predictions. Averaging over expectational models lends itself to the intuitive interpretation that subjects have changing degrees of confidence in the different models.

Our model can be formalized following the general learning algorithm proposed by Arthur (1993). Let $\vec{m}_t = \{m_t^{RE}, m_t^{ME}, m_t^{TR}\}$ be a vector of the individual model predictions under rational, myopic and trend-following expectations. These models yield the same individual predictions as described in Section 2. Let $\vec{c}_t = \{c_t^{RE}, c_t^{ME}, c_t^{TR}\}$ be a vector of *confidence* assigned to each model m before the start of period t .¹⁶ Further, let C_t be the current sum of these confidence measures (component sum of \vec{c}_t). We assume the sum of confidence measures to be constant over time, i.e. we set $C_t = C$. This implies that the parameter C controls the speed of learning or responsiveness to new price signals, which is decreasing in C .

Before the start of each period t , subjects:

1. Calculate the vector of weights \vec{w}_t associated with each model. That is $\vec{w}_t = \vec{c}_t/C$.
2. Form a belief b_t^{ALM} as the weighted average of individual model predictions. That is, $b_t^{ALM} = \vec{w}_t \times \vec{m}_t$
3. Observe the deviation of each individual model, j , from the actual price in period t , $d_t^j = |m_t^j - p_t|$.
4. Update the confidence vector. That is, $\vec{c}_{t+1} = \vec{c}_t + \vec{\beta}_t$, where $\vec{\beta}_t = \{(1 + d_t^{RE})^{-1}, (1 + d_t^{ME})^{-1}, (1 + d_t^{TR})^{-1}\}$. Accordingly, subjects place greater confidence on models with lower deviation d .
5. Renormalize the confidence vector. The new confidence placed in each model in the

¹⁶Like Rational Expectations, our ALM assumes homogeneous beliefs for simplicity, even though we observe heterogeneous beliefs in our markets. The ALM could be extended to allow for heterogeneous beliefs, for example by assuming that subjects initially place different weights on the three models depending on their cognitive ability or their academic background.

next period is then defined as,

$$\vec{c}_{i,t+1} = \frac{C}{C + B_t}(\vec{c}_{i,t} + \beta_t) \quad (12)$$

where B_t is the component sum of β_t .¹⁷

We estimate this model on participants' beliefs for various learning speeds and compare the goodness of fit to our previous models. The estimates are reported in Table 9. The ALM describes describes beliefs better than all other models in both first and second markets, as well as Bull and Bear markets. As we increase the speed of learning (by lowering the parameter C) the model's accuracy improves further. This outcome is not surprising, because the ALM is flexible and allows subjects to revise their beliefs in light of market dynamics, unlike the other models.

Table 9: RMSEs of individual beliefs and model predictions

	RE	Myopic	Trend	ALM (C=0.1)	ALM (C=1)	ALM (C=10)
1st Market	120.06	127.31	98.36	77.64	86.19	89.54
2nd Market	70.76	97.48	77.18	52.43	54.74	56.48
Bull	112.31	116.12	77.4	67.34	74.56	77.27
Bear	78.07	108.02	98.74	62.74	66.39	68.75

This table reports the observed deviations of beliefs from the predictions implied by the following models: Rational Expectations Model (RE), Myopic Model (Myopic), Trend Model (Trend), and Adaptive Learning Model (ALM). We calculate the root mean squared errors (RMSE) at the individual market level (16 markets per cell) and present the average over treatments. For the ALM we assume equal initial confidence in the models in period 3. Columns 4-6 vary the learning parameter C. A greater value of C implies a lower speed of learning.

More interestingly, however, our ALM allows us to illustrate how subjects' confidence in Rational Expectations increases as the experiment progresses. For this purpose we plot the

¹⁷A numerical example helps to illustrate the dynamics of our learning model: Assume that subjects are equally confident in each model at the beginning of period 3, after a positive announcement shock in period 2. That is, $\vec{c}_{t=3} = \{1, 1, 1\}$ and $C = 3$. Assume further that the Rational Expectations, Myopic and Trend-following models yielded the following predictions $\vec{m}_{t=3} = \{460, 310, 370\}$. Given equal confidence, subjects will assign a weight of $1/3$ to each model. Thus, under Adaptive Learning subjects predict $b_{t=3}^{ALM} = 380$ for period 3.

Next, suppose that the actual trading price in period 3 was 400. The subject calculates deviations $d_{t=3} = \{60, 90, 30\}$. These deviations then result in an updated confidence vector $\vec{c}_{t=4} = \{0.996, 0.991, 1.012\}$ with new weights $\vec{w}_{t=4} = \{0.332, 0.330, 0.337\}$. In this example, the Trend model was most accurate in period 3 and therefore gains relative weight in prediction. If we increase the speed of learning (decrease C to 0.06 for example) the weights would react more strongly: $\vec{w}_{t=4} = \{0.304, 0.259, 0.437\}$.

average weights that subjects place on Rational Expectations, Myopic and Trend-following expectations (Figure 6) across treatments. A number of patterns emerge. In all treatments, subjects start placing the highest weight on Rational Expectations right after the announcement shock at the beginning of period 3, and RE dominates thereafter except for one brief interruption. The dominance of Rational Expectations is most pronounced in second markets and in Bear markets, which is in line with our previous findings. Weights converge towards the end of the market, as the predictions from the Rational Expectations and Myopic models coincide more closely at the end of life of the asset.

This modeling exercise reinforces our view that Rational Expectations does describe observed behavior in our experimental setting well. At the same time, when we relax the assumption of an instantaneous reaction to the announcement shock and allow subjects to revise beliefs through our Adaptive Learning Model, accuracy improves. Model fit might improve even further were we to extend the model to capture heterogeneous beliefs or different reactions to positive and negative announcement shocks. While these extensions are beyond the scope of this paper, our model would be flexible enough to incorporate them.

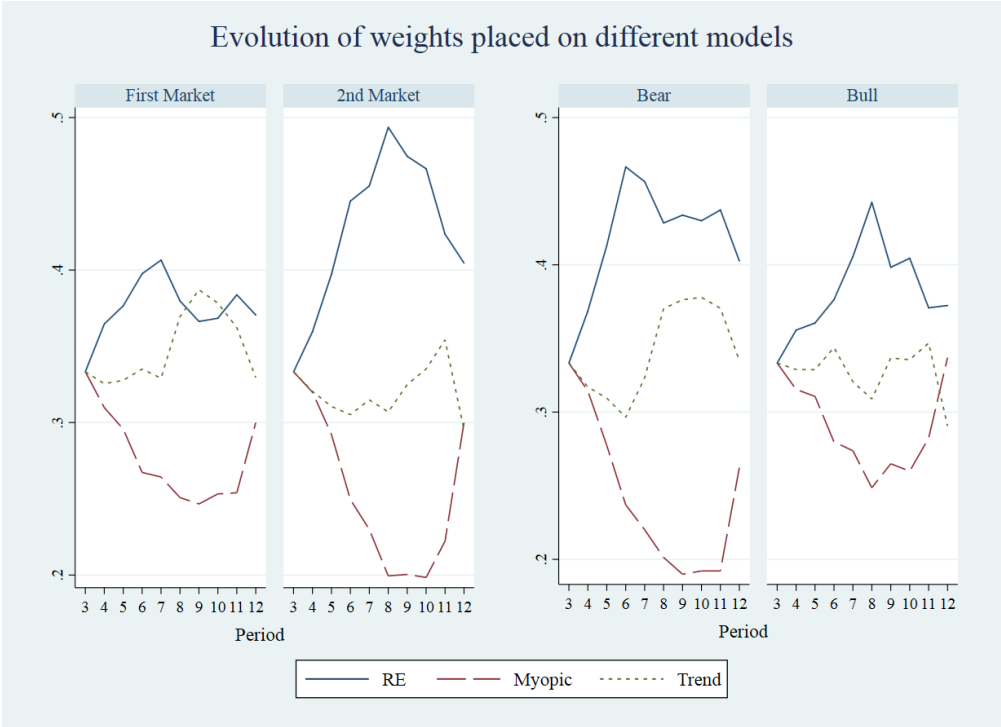


Fig. 6. Evolution of weights placed on different models
 The figure shows the weights placed on the three models of expectation formation over time within the adaptive learning model. The value of the parameter C was set to 1.

7. Conclusion

In experimental finance, the view that long-lived asset markets tend to exhibit mispricing is widely held. This intuition is based on the behavior of the dominant paradigm in experimental finance, introduced by Smith et al. (1988). The motivation for our study was to construct a simple asset market with an alternative structure, and to evaluate whether it exhibits similarly severe mispricing. In our setting, the assets trading in the market are shares in a firm whose earnings follow a random process with either positive or negative expected value. All earnings are retained until the end of the market. The earnings regime is announced early in the life of the asset. This announcement causes an immediate and pronounced shift in the firm's expected future earnings. The timing of the release of new information allows the predictions of different models of price and expectation formation to be distinguished. The delayed onset of the price trend allows a reallocation of holdings of assets and cash among traders before the principal event of interest.

Our price data provide considerable support for the Rational Expectations Model. A claim that the data are *close* to a model's prediction is always subjective, since there are individual cases where deviations from the model's predictions, in terms of prices and beliefs are quite large. In our view, a reasonable standard is to use the related literature and the performance of comparable models as a basis of comparison. By this standard, we feel that the claim that the RE model is supported in the data is justified. The analysis in section 5 shows that that prices are closer to RE levels than in related studies. Furthermore, the RE model is more accurate than two alternative models we have proposed, the Trend and Myopic models. If we allow for traders to switch between Rational Expectations and other models, we can achieve additional explanatory power. However, the relatively good performance of the model incorporating switching is perhaps not surprising, and more stringent tests of this model would be when it is taken to new data and compared to other flexible expectational models.

In addition to exploring aggregate patterns, we analyze belief formation at the individual level. In particular, we study the relationship between subjects' expectations and their cognitive ability, as measured with the Cognitive Reflection Test. We find that those scoring more highly on the test adhere more closely to the Rational Expectations predictions and less closely to the Trend model. We also find that relatively sophisticated subjects achieve higher earnings, although their advantage decreases from the first to the second market, as there is less mispricing to exploit. Unlike previous studies, mispricing in our markets is uncorrelated with a cohort's average CRT score for the range of cognitive ability we observe among our participants. We interpret this as evidence that our design is robust, in achieving rational expectations pricing, to the introduction of relatively unsophisticated traders, at

least to some extent.

In our view, our results have implications for the study of expectation formation and market efficiency in experimental asset markets. Our markets tend to exhibit prices close to the Rational Expectations level. We believe that our design is well-suited to provide further experimental evidence regarding the debate about whether asset prices under- or overreact to news shocks (Fama, 1998). In our data, we do not observe systematic long-term deviations from Rational Expectations, only transitory departures.

Why does Rational Expectations predict behavior relatively well here? It is difficult to isolate the precise factors underlying the difference between the degree of adherence of prices to Rational Expectations in our experiment compared to others. However, (1) we believe that having the fundamental value be approximately constant over most of the lifespan of the asset is conducive to better pricing. Constant fundamental value trajectories have been shown to be relatively conducive to pricing at fundamentals by Noussair et al. (2001), and Kirchler et al. (2012), and it seems that a similar effect is operating in our environment. (2) The news shock in our setting is salient and occurs to the *trend* of the fundamental value process, which directly affects the level of all remaining earnings announcements. This might induce forward-looking behavior more than in other experimental settings, where earnings announcements are typically incremental. (3) There is a period of stationarity at the beginning before the trend sets in, allowing portfolio rebalancing and sophisticated traders to accumulate more weight in the market before the trend begins. This means a greater share of trades is done for arbitrage rather than rebalancing, and is done by relatively sophisticated traders, during the trend phase when it is more difficult to form expectations. Future experiments could consider these possibilities and try to identify the crucial factors behind the strong performance of Rational Expectations.

Follow-up work could also consider the robustness of the tendency to price at Rational Expectations in extensions of our setting. Known contributors to bubble formation, such as greater cash endowments, longer trading horizons, or other forces that reduce the prominence of the fundamental anchor and induce speculation (Hirota and Sunder, 2007; Lahav, 2011; Hirota et al., 2015) might cause bubbles to form. On the other hand, if mispricing is not commonly observed under these more challenging conditions, than the RE model would be even more strongly supported than it is here.

References

- Abarbanell, J. and Bernard, V. (2000). Is the U.S. Stock Market Myopic. *Journal of Accounting Research*, 38(2):221–242.
- Adam, K., Marcet, A., and Nicolini, J. P. (2016). Stock market volatility and learning. *The Journal of Finance*, 71(1):33–82.
- Arthur, W. B. (1993). On designing economic agents that behave like human agents. *Journal of Evolutionary Economics*, 3(1):1–22.
- Ball, R. and Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, 6(2):159–178.
- Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.
- Bernard, V. L. and Thomas, J. K. (1989). Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium? *Journal of Accounting Research*, 27(3):1–36.
- Bosch-Rosa, C., Meissner, T., and Bosch-Domènech, A. (2018). Cognitive bubbles. *Experimental Economics*, 21(1):132–153.
- Brav, A., Jiang, W., and Tian, X. (2014). Shareholder Power and Corporate Innovation: Evidence from Hedge Fund Activism.
- Breaban, A. and Noussair, C. N. (2015). Trader characteristics and fundamental value trajectories in an asset market experiment. *Journal of Behavioral and Experimental Finance*, 8:1–17.
- Bruguier, A. J., Quartz, S. R., and Bossaerts, P. (2010). Exploring the nature of "trader intuition". *Journal of Finance*, 65(5):1703–1723.
- Caginalp, G., Hao, L., Porter, D., and Smith, V. L. (2008). Asset market reactions to news: an experimental study. *Chapman University, Economic Science Institute Working Papers*, pages 1–51.
- Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, Princeton, NJ, USA.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikenstein, E., Holzmeister, F., Imai, T.,

- Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280):1433–1437.
- Carle, T. A., Lahav, Y., Neugebauer, T., and Noussair, C. N. (2016). Heterogeneity of beliefs and trade in experimental asset markets. *Working Paper*.
- Carpenter, J., Graham, M., and Wolf, J. (2013). Cognitive ability and strategic sophistication. *Games and Economic Behavior*, 80:115–130.
- Charness, G. and Neugebauer, T. (2018). A test of the modigliani-miller invariance theorem and arbitrage in experimental asset markets. *The Journal of Finance*, forthcoming.
- Corgnet, B., DeSantis, M., and Porter, D. (2018). What makes a good trader? on the role of intuition and reflection on trader performance. *The Journal of Finance*, 73(3):1113–1137.
- Corgnet, B., Herna, R., Kujal, P., and Porter, D. (2015). The Effect of Earned Versus House Money on Price Bubble Formation in Experimental Asset Markets. *Review of Finance*, 19(4):1455–1488.
- Cueva, C. and Rustichini, A. (2015). Is financial instability male-driven? gender and cognitive skills in experimental asset markets. *Journal of Economic Behavior & Organization*, 119:330–344.
- Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance*, 53(6):1839–1885.
- Duffy, J. (2006). Agent-Based Models and Human Subject Experiments. In *Handbook of Computational Economics Vol. 2*, chapter 19, pages 950–1011.
- Dufwenberg, M., Lindqvist, T., and Moore, E. (2005). Bubbles and Experience: An Experiment. *American Economic Review*, 95(5):1731–1737.
- Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, pages 848–881.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):28–30.
- Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3):283–306.

- Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. (1969). The Adjustment of Stock Prices to New Information. *International Economic Review*, 10(1):1–21.
- Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Readymade Economics Experiments. *Experimental Economics*, 10(2):171–178.
- Forsythe, R., Palfrey, T. R., and Plott, C. R. (1982). Asset Valuation in an Experimental Market. *Econometrica*, 50(3):537–567.
- Frazzini, A. (2006). The disposition effect and underreaction to news. *Journal of Finance*, 61(4):2017–2046.
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4):25–42.
- Friedman, D., Harrison, G. W., and Salmon, J. W. (1984). The Informational Efficiency of Experimental Asset Markets. *Journal of Political Economy*, 92(3):349–408.
- Greiner, B. (2003). An Online Recruitment System for Economic Experiments. *Forschung und wissenschaftliches Rechnen*, 63:79–93.
- Grinblatt, M. and Han, B. (2005). Prospect theory, mental accounting, and momentum. *Journal of Financial Economics*, 78(2):311–339.
- Grossman, S. (1976). On the efficiency of competitive stock markets where trades have diverse information. *The Journal of Finance*, 31(2):547–585.
- Grossman, S. and Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408.
- Hanaki, N., Akiyama, E., Funaki, Y., Ishikawa, R., et al. (2017). Diversity in cognitive ability enlarges mispricing in experimental asset markets. *GREDEG Working Paper*.
- Haruvy, E., Lahav, Y., and Noussair, C. N. (2007). Traders’ expectations in asset markets: Experimental evidence. *American Economic Review*, 97(5):1901–1920.
- Haruvy, E. and Noussair, C. N. (2006). The Effect of Short Selling on Bubbles. *Journal of Finance*, 59(3):1119–1157.
- Hefti, A., Heinke, S., and Schneider, F. (2016). Mental Capabilities, Trading Styles, and Asset Market Bubbles: Theory and Experiment. *Working Paper*.

- Hirota, S., Huber, J., Stöckl, T., and Sunder, S. (2015). Investment Horizons and Price Indeterminacy in Financial Markets. *Working Paper*.
- Hirota, S. and Sunder, S. (2007). Price bubbles sans dividend anchors: Evidence from laboratory stock markets. *Journal of Economic Dynamics and Control*, 31(6):1875–1909.
- Holt, C. and Laury, S. (2002). Risk aversion and incentive effects. *The American Economic Review*, 92(5):1644–1655.
- Hommel, 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.
- Hong, H., Stein, J., and Yu, J. (2007). Simple Forecasts and Paradigm Shifts. *The Journal of Finance*, 62(3):1207–1242.
- Hong, H. and Stein, J. C. (1999). A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *The Journal of Finance*, 54(6):2143–2184.
- Hussam, R. N., Porter, D., and Smith, V. L. (2008). Thar She Blows: Can Bubbles Be Rekindled with Experienced Subjects? *American Economic Review*, 98(3):924–937.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8):0696–0701.
- Jegadeesh, N. and Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1):65–91.
- King, R. R., Smith, V. L., Williams, A. W., and Van Boening, M. (1993). The robustness of bubbles and crashes in experimental stock markets. *Nonlinear dynamics and evolutionary economics*, pages 183–200.
- Kirchler, M. (2009). Underreaction to fundamental information and asymmetry in mispricing between bullish and bearish markets. An experimental study. *Journal of Economic Dynamics and Control*, 33(2):491–506.
- Kirchler, M., Huber, J., and Stockl, T. (2012). Thar She Bursts-Reducing Confusion Reduces Bubbles. *American Economic Review*, 102(2):865–883.
- Kuhnen, C. M. (2015). Asymmetric Learning from Financial Information. *Journal of Finance*, 70(5):2029–2062.
- Lahav, Y. (2011). Price patterns in experimental asset markets with long horizon. *Journal of Behavioral Finance*, 12(1):20–28.

- Lei, V., Noussair, C. N., and Plott, C. R. (2001). Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality. *Econometrica*, 69(4):831–859.
- Lin, S. and Rassenti, S. (2012). Are under- and over-reaction the same matter? Experimental evidence. *Journal of Economic Behavior and Organization*, 84(1):39–61.
- Maniadis, Z., Tufano, F., and List, J. A. (2014). One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *American Economic Review*, 104(1):277–290.
- Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica*, 29(3):315–335.
- Noussair, C., Ruffieux, B., and Robin, S. (2001). Price Bubbles in Laboratory Asset Markets with Constant Fundamental Values. *Experimental Economics*, 4(1):87–105.
- Noussair, C. N., Trautmann, S. T., and Van De Kuilen, G. (2014). Higher Order Risk Attitudes, Demographics, and Financial Decisions. *Review of Economic Studies*, 81:325–355.
- Noussair, C. N. and Tucker, S. (2016). Cash inflows and bubbles in asset markets with constant fundamental values. *Economic Inquiry*, 54(3):1796 – 1606.
- Oechssler, J. (2010). Searching beyond the lamppost: Let's focus on economically relevant questions. *Journal of Economic Behavior and Organization*, 73(1):65–67.
- Palan, S. (2013). A Review of bubbles and crashes in experimental asset markets. *Journal of Economic Surveys*, 27(3):570–588.
- Palan, S. (2015). GIMS — Software for asset market experiments. *Journal of Behavioral and Experimental Finance*, 5:1–14.
- Plott, C. R. and Sunder, S. (1982). Efficiency of Experimental Security Markets with Insider Information: An Application of Rational-Expectations Models. *Journal of Political Economy*, 90(4):663–698.
- Plott, C. R. and Sunder, S. (1988). Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56(5):1085–1118.
- Powell, O. (2016). Numeraire independence and the measurement of mispricing in experimental asset markets. *Journal of Behavioral and Experimental Finance*, 9(March 2016):56 – 62.

- Powell, O. and Shestakova, N. (2016). Experimental asset markets: A survey of recent developments. *Journal of Behavioral and Experimental Finance*, 12:14–22.
- Shiller, R. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *The American Economic Review*, 71(3):421 – 436.
- Smith, V. L. (1962). An Experimental Study of Competitive Market Behavior. *Journal of Political Economy*, 70(2):111–137.
- Smith, V. L. (2002). Method in Experiment: Rhetoric and Reality. *Experimental Economics*, 5(2):91–110.
- 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.
- Stöckl, T., Huber, J., and Kirchler, M. (2010). Bubble measures in experimental asset markets. *Experimental Economics*, 13(3):284–298.
- Stöckl, T., Huber, J., and Kirchler, M. (2014). Multi-period experimental asset markets with distinct fundamental value regimes. *Experimental Economics*, 18(2):314–334.
- Sunder, S. (1992). Market for Information : Experimental Evidence. *Econometrica*, 60(3):667–695.
- Thaler, R. H., Tversky, A., Kahneman, D., and Schwartz, A. (1997). The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test. *The Quarterly Journal of Economics*, 112(2):647–661.
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2):147–168.
- Walker, E. and Nowacki, A. S. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine*, 26(2):192–196.
- Weber, M. and Welfens, F. (2007). How do Markets React to Fundamental Shocks? An Experimental Analysis on Underreaction and Momentum. *Working Paper*.

Appendix A. Additional Analyses

In this appendix, we state and evaluate two additional hypotheses regarding the relationship between subject characteristics and their behavior. The first is the relationship between risk aversion and asset holdings. Because the asset is a lottery over future payoffs, we would expect subjects who are more risk-averse than their peers to decrease their share-holdings over time. This is stated as our fifth hypothesis.

Hypothesis 5. *Subjects who are relatively more (less) risk-averse than the cohort average decrease (increase) their holdings of the risky asset over time.*

The next hypothesis concerns the relative performance of more versus less sophisticated participants. We use the CRT score of an individual as a measure of her sophistication. A number of studies have found that individuals with higher CRT scores earn more in asset market experiments than those with lower scores (Breaban and Noussair, 2015; Bosch-Rosa et al., 2018; Corgnet et al., 2015; Charness and Neugebauer, 2018). Those scoring higher on other measures of cognitive ability, like the Hit-15 game and the pattern task, also earn greater trading profits (Cueva and Rustichini, 2015). We hypothesize that a similar correlation would be observed here as well.

Hypothesis 6. *Subjects with higher cognitive ability earn more than those with lower cognitive ability.*

Figure 7 plots changes in asset holdings over time for different groups of subjects in both first and second markets. The right-hand panel in Figure 7 shows the share of assets held by subjects with below-median degrees of risk aversion, where risk aversion is measured by the number of risky choices made in the protocol by Noussair et al. (2014). The figure shows that those traders who have relatively low levels of risk aversion increase their holdings of the asset over time in both the first and second markets. As described in Result 11, subjects with lower risk aversion are net buyers of the asset in both markets.

Result 11: Subjects who are relatively more (less) risk-averse decrease (increase) their holdings on average.

The left panel plots the total share of assets held by subjects with CRT scores above 4. In first markets, sophisticated subjects increase their asset holdings to more than 60% of all shares by the end of the second period. This gives them significant power to influence market prices and further explains why prices react relatively efficiently to news in our design: sophisticated subjects hold most assets when the news shock arrives and will only dispose of their assets at prices that marginally exceed fundamental values. In second markets, when all subjects have gained some experience, sophisticated subjects no longer play such a dominant role, although they are still net buyers on average.

To test hypothesis 6, we regress subjects' trading profits on their cognitive ability, program of study, and risk attitudes. Their risk attitudes are measured by the number of risky and imprudent choices they make in the tasks administered at the beginning of the sessions.

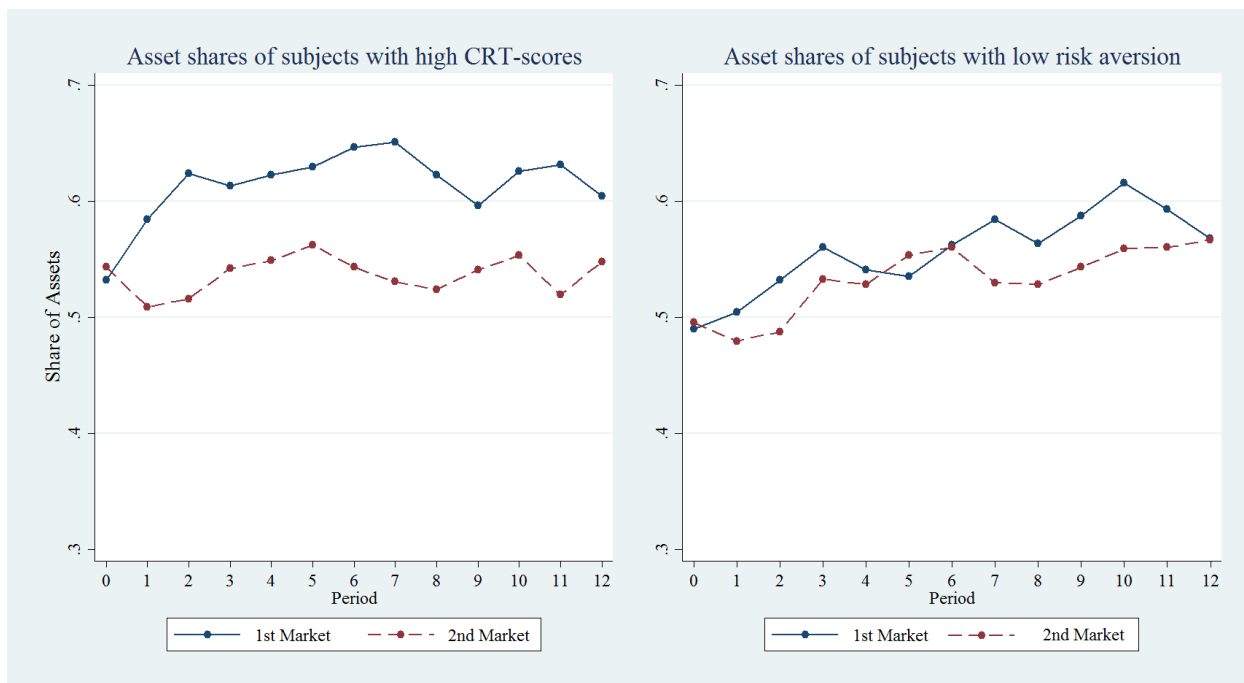


Fig. 7. Changes in Asset Holdings conditional on CRT and Risk Aversion

This figure plots changes in the shares of assets held by different groups of subjects in first and second markets. The left panel of this figure compares subjects with CRT scores above the median (blue solid line) against subjects with CRT scores below the median (red dashed line). The right panel compares subjects with high risk aversion (blue solid line) against subjects with low risk aversion (red dashed line). The degree of risk aversion is measured by the number of risky choices made in the protocol first employed by Noussair et al. (2014).

Trading profits are defined as the difference between initial and final wealth.

$$\begin{aligned}
 Profit_i = & \beta_0 + \beta_1 \times CRT_i + \beta_2 \times Econ_i + \beta_3 \times Bull \\
 & + \beta_4 \times RiskSeekingness_i + \beta_5 \times Imprudence_i + \epsilon_i
 \end{aligned}
 \tag{13}$$

The determinants of trading profits are shown in Table 10. On the one hand, they reflect the finding reported earlier that subjects with higher cognitive ability adhere more closely to Rational Expectations and form more precise predictions. On the other hand, the results reflect the fact that price discovery improves with repetition. More sophisticated subjects earn on average higher profits in the first markets, but the premium they earn decreases in the second markets. This is because there is less mispricing to exploit. The findings are reported as Result 12.

Result 12: Subjects with higher CRT scores earn more than those with lower scores. The relationship is weaker in the second market than in the first.

The better performance of more sophisticated subjects raises the question of which trad-

Table 10: Determinants of trading profits

	1st Markets Profit	2nd Markets Profit	1st Markets Profit	2nd Markets Profit
CRT	714.8*** (3.53)	210.2*** (3.24)	675.0*** (3.12)	200.1*** (3.15)
Econ	987.3 (1.24)	270.9 (1.32)	871.9 (0.94)	189.2 (0.81)
Bull	110.3 (0.38)	170.7* (1.95)	206.1 (0.63)	165.0 (1.48)
Risk Seekingness			215.8 (0.97)	114.7 (1.12)
Imprudence			-204.5 (-1.13)	12.90 (0.13)
Constant	-600.4 (-1.35)	-238.1* (-2.03)	-584.9 (-1.19)	-190.3 (-1.26)
Observations	126	126	126	126
R^2	0.141	0.131	0.147	0.143

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents the cross-sectional regression of individual trading profits on a number of covariates. Trading profits are defined as the difference between initial endowment and final wealth. The regression is conducted separately for the performance in the first and the second market. CRT is the individual's CRT score. CRT scores are centralized and unit changes can be interpreted as the deviations from the CRT sample mean of 4.53. Econ indicates whether or not the subject has an Economics/Business background of study. Risk Seekingness and Imprudence indicate the number of risky and imprudent choices made in the lotteries preceding the markets and are centered at their means of 2.93 and 0.93, respectively. Bull is a dummy, equaling 1 for a Bull and 0 for a Bear markets. Observations from Session 1 and 2 were excluded, as CRT was measured on a 3-item scale in these sessions. Standard errors are clustered at the market-level.

ing strategies allow them to achieve these superior earnings. For this purpose, we look closer at individual purchases and sales. If sophisticated subjects engage in rational arbitrage, they should be buying in periods during which the prevailing trading price lies below the fundamental value and should be selling in periods during which it lies above. Similarly to Corgnet et al. (2015), Table 11 regresses subjects' net purchases (buys - sells) per period on the relative deviation (RD) in that period. The regression includes a dummy that equals one for subjects with a CRT score above the median of 4, and an interaction term between high CRT score and RD. If this interaction term is significantly negative, it means that those taking more profitable positions are those with higher CRT scores. The results are presented separately for first and second markets. The proposed relationship, embodied in the interaction term, is statistically significant only in first markets when there is mispricing to exploit, but not in second markets. The significant coefficient reveals that high CRT individuals, who are presumably more sophisticated, are more likely to purchase at prices that are lower relative to fundamentals.¹⁸

Result 13: Arbitrage activity is stronger in first than in second markets.

¹⁸The left-hand panel of Figure 7 indicates that traders with CRT scores above the median hold a clear majority of the units in Market 1. However, in Market 2, they hold approximately 50 percent of the units, an approximately equal share as those traders with below-median CRT scores.

Table 11: Net purchases conditional on CRT-score and mispricing

	1st Market Net Purchases	2nd Market Net Purchases
CRT^{high}	-0.0222 (-0.14)	0.141 (1.16)
RD_t	1.266*** (2.63)	0.657 (1.10)
$CRT^{high} \times RD_t$	-2.013** (-2.56)	-0.979 (-1.08)
Constant	0.0115 (0.12)	-0.0874 (-1.22)
Observations	1584	1548
R^2	0.00525	0.000869
$RD_t + CRT^{high} \times RD_t$	-0.747**	0.322

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents the results of a random-effects regression of subjects' net purchases (buys - sells) per period on a number of covariates, separately for first and second markets. CRT^{high} is a dummy variable that equals one if a subject's CRT score is above the mean of 4. RD_t is the relative mispricing in a period and is defined as: $RD_t = \tilde{P}_t - RE_t / RE_t$ where \tilde{P}_t is the average trading price and RE_t is the rational expectations price in a period. Standard errors are clustered at the market-level. The number of observations refers to the number of subject-period observations.

Appendix B. Efficiency by market and period

Table 12: Efficiency statistics at the individual market level

Panel A								Panel B				
Session	Treatment	RAD	RD	GAD	GD	Turnover	Price Dispersion	Period	Treatment	RAD	Turnover	Price Dispersion
1	bull	26.1%	-24.0%	44.8%	-29.6%	9.7%	17.9%	1		25.0%	0.56	35.9%
1	bear	18.1%	2.9%	20.3%	-1.1%	5.8%	9.8%	2		20.2%	0.41	16.1%
2	bear	10.6%	-9.5%	9.7%	-7.7%	23.7%	9.9%	3		30.4%	0.44	21.6%
2	bull	3.4%	-3.3%	3.9%	-3.6%	7.1%	9.7%	4		24.1%	0.37	10.6%
3	bear	13.4%	-1.6%	12.3%	0.9%	12.6%	5.7%	5		16.1%	0.34	12.3%
3	bull	3.2%	-1.9%	3.4%	-2.0%	12.0%	2.6%	6		13.4%	0.40	10.3%
4	bull	3.9%	0.2%	4.4%	-0.5%	30.4%	6.4%	7	Bull	14.2%	0.26	6.8%
4	bear	9.1%	6.3%	8.2%	4.6%	32.7%	8.1%	8		13.7%	0.28	9.3%
5	bear	22.8%	15.4%	24.7%	18.8%	48.0%	29.7%	9		16.3%	0.22	11.2%
5	bull	7.7%	-0.1%	7.9%	-0.2%	41.5%	38.1%	10		13.6%	0.21	24.9%
6	bull	15.6%	-13.7%	21.3%	-16.2%	50.7%	31.0%	11		5.9%	0.24	4.6%
6	bear	13.4%	11.8%	10.7%	8.4%	28.3%	17.0%	12		7.1%	0.22	4.8%
7	bear	28.9%	-7.6%	30.4%	-3.8%	22.8%	29.2%	1		50.4%	0.48	33.1%
7	bull	8.8%	0.3%	12.2%	-3.3%	8.5%	7.8%	2		35.9%	0.34	22.1%
8	bull	9.8%	1.9%	10.7%	1.8%	20.1%	34.4%	3		28.5%	0.47	18.2%
8	bear	14.7%	14.7%	12.3%	12.3%	24.2%	6.9%	4		16.9%	0.37	10.8%
9	bear	13.7%	12.9%	11.1%	10.2%	17.6%	3.8%	5		13.9%	0.33	8.9%
9	bull	8.1%	-5.4%	8.1%	-4.8%	14.9%	1.2%	6	Bear	10.3%	0.40	7.6%
10	bear	13.9%	-13.7%	16.2%	-13.8%	32.8%	18.1%	7		12.1%	0.32	11.7%
10	bull	9.6%	-6.0%	10.8%	-5.3%	13.6%	13.8%	8		8.3%	0.24	4.7%
11	bear	44.4%	-33.5%	108.7%	-46.9%	53.8%	25.3%	9		6.4%	0.28	17.0%
11	bull	5.0%	0.1%	6.2%	-0.3%	11.1%	10.4%	10		6.9%	0.31	14.2%
12	bear	7.4%	-4.5%	8.3%	-5.2%	30.6%	17.6%	11		4.5%	0.25	6.0%
12	bull	6.2%	5.9%	7.5%	7.2%	17.6%	2.5%	12		6.3%	0.18	3.8%
13	bull	49.6%	19.8%	62.2%	-0.5%	43.8%	7.7%					
13	bull	22.8%	19.2%	23.1%	19.0%	28.4%	3.9%					
14	bull	40.5%	-8.9%	69.6%	-24.7%	19.6%	36.9%					
14	bear	20.6%	7.5%	20.9%	5.7%	16.8%	12.8%					
15	bull	35.0%	-31.0%	59.2%	-35.0%	53.7%	7.3%					
15	bear	26.5%	24.7%	25.0%	23.2%	26.3%	10.9%					
16	bull	17.9%	-17.6%	35.2%	-25.9%	28.2%	23.7%					
16	bear	2.0%	-1.1%	1.8%	-0.9%	6.9%	0.7%					
Panel C												
Mean bull		16.6%	-3.7%	23.9%	-7.4%	0.25	15.9%					
Mean bear		16.7%	1.2%	20.5%	0.0%	0.25	12.9%					
Mean 1st		22.1%	-7.2%	33.1%	-11.1%	0.31	19.0%					
Mean 2nd		11.2%	4.7%	11.4%	3.7%	0.18	9.8%					
Median bull		9.7%	-1.0%	11.5%	-2.6%	0.20	10.0%					
Median bear		13.8%	0.9%	12.3%	0.0%	0.24	10.4%					
Median 1st		16.7%	-8.3%	23.0%	-6.4%	0.29	18.0%					
Median 2nd		8.9%	1.6%	9.4%	-0.2%	0.16	8.9%					

This table provides mispricing measures for all markets. In Panel A, the results for all individual markets are displayed. In Panel B, the averages over all periods in Bull and Bear markets are shown, providing insights about the dynamics. In Panel C, the mean and median estimates for our different treatments are shown. The mispricing measures calculated as follows: For each market m Relative Absolute Deviation (RAD) = $\frac{1}{T} \sum_{t=1}^T \frac{|\overline{P}_t - RE_t|}{\overline{RE}_m}$, where \overline{P}_t is the average price in a period and RE_t is the rational expectations value in that period and \overline{RE}_m is the average rational expectations value in that market. RD is the Relative Deviation, defined as $\frac{1}{T} \sum_{t=1}^T \frac{\overline{P}_t - RE_t}{\overline{RE}_m}$. Geometric Absolute Deviation (GAD, Powell (2016)) = $\exp(\frac{1}{T} \sum_{t=1}^T |\ln(\frac{\overline{P}_t}{RE_t})|)$. GD is the analogous geometric deviation. Turnover is defined as the total trading volume over the course of the market, divided by the total number of shares held by all traders. Price Dispersion measures the price volatility within periods and is defined as: Price Dispersion = $\frac{1}{T} \sum_{t=1}^T \sqrt{\frac{1}{N} \sum_{n=1}^N (RET_n - \overline{RET}_t)^2}$ where $RET_n = \ln(\frac{P_n}{P_{n-1}})$.

Appendix C. Experimental Instructions

Introduction. Welcome to this experiment. Please refrain from talking to the other participants for the duration of this study. This study will last for two hours and your payment at the end will depend upon your performance in the study. We will start with an explanation of the experiment. If you have any questions regarding the instructions, please do not hesitate to ask.

In this experiment, you will have the opportunity to participate in a stock market. You will participate in two stock markets sequentially, and the first market is independent from the second. In each market, you can trade shares of a firm with the other participants on a computerized trading platform. As you can see, nine subjects will be participating in each market. Each market lasts for 12 trading periods, each of which is 2 minutes long. For each market, you will receive a starting budget, consisting of a number of shares of the company and some cash.

Information on the firm. You are given some information on the firm, which is relevant for trading. First of all, the firm does not pay out any dividends throughout the market. However, the firm makes profits or losses each period and announces them at the end of each period. These profits or losses are retained until the end of a market and then paid out. Thus, if you hold a share of the firm at the end of period 12, you will receive the total earnings that the firm has accumulated until then. In each market, the firm will start with an initial amount of earnings, which is announced at the beginning of a market.

The level of the firm's period earnings depends on the firm's **business conditions**. The business conditions may be *good*, *bad* or *neutral*. You all have the following information:

- Period 1-2: The business conditions will be *neutral*. Each period, the profits or losses per share may be +10 Taler or -10 Taler with equal probability (50% - 50%).
- Period 3-10: The business conditions will be either *good* or *bad*. Both possibilities are **equally likely**. Either the good or the bad business conditions last for the entire 8 periods and it is not possible for business conditions to switch during these periods.
If the business conditions are *good*: Period earnings per share may be +10 Taler or +30 Taler per period with equal probability (50% - 50%).
If the business conditions are *bad*: Period earnings per share may be -10 Taler or -30 Taler per period with equal probability (50% - 50%).
- Period 11 - 12: The business conditions will be *neutral* again. Each period, the profits or losses per share may be +10 Taler or -10 Taler with equal probability (50% - 50%).

Important: Before the beginning of period 3, the firm announces whether the good or the bad business conditions prevail for the subsequent 8 periods. The business conditions are unpredictable and are independent of what happened in periods 1 and 2. This information is again summarized in the following figure (*Figure 2 in the paper.*)

Forecasting. In addition to the possibility of trading shares, you will be asked to make forecasts. Before the beginning of each trading period, you will be asked to make a forecast about the average (volume-weighted) trading price in the upcoming period. Additionally, you will be asked how certain you are about your forecast. The following figure displays the screen used for entering your forecasts. On the right, you enter your forecast and your degree of certainty and on the left, you see the history of your forecasts and the actual prices.

You will be paid according to the accuracy of your forecast. For each forecast that lies within a 10% range of the actual forecast, you receive 0.20 Euro. At the end of the experiment, you will be paid according to your forecast accuracy in one of the markets.

The screenshot shows a forecasting interface. At the top, it indicates 'Periode 2 von 12' and 'Verbleibende Zeit (Sekunden): 19'. The interface is split into two main sections.

Left Section: Ihre Vorhersagen in den letzten Perioden

Periode	Vorhersage	Tatsächlicher Durchschn. Handelspreis	Anzahl präziser Vorhersagen
1	230.00	160.00	0

Three blue boxes highlight the following data points from the table:

- Price forecast in previous periods (230.00)
- Average price in previous periods (160.00)
- Number of accurate forecasts (0)

Right Section: Ihre Vorhersage für die nächste Periode

This section contains a form for entering a forecast and a certainty level. A blue box highlights the 'Price forecast for upcoming period' input field. Below it is a label 'Ihre Vorhersage für den durchschnittlichen Handelspreis in der nächsten Periode' followed by an empty input field. To the right of this field is a red 'OK' button.

Below the input field is a question: 'Wie sicher sind Sie sich bei Ihrer Vorhersage?' followed by a scale from 'sehr unsicher' to 'sehr sicher' with five empty circles. A blue box highlights the 'Degree of certainty about price forecast (very unsure – very sure)' label.

Functioning of the trading platform. Next, we turn to the functioning of the trading platform. Before the start of the experiment, you will have the chance to play three practice rounds to become familiar with the platform.

The price at which you can trade a share is determined by your actions and those of the other participants. There are two ways in which you can trade a share:

- You can place a buy or sell order yourself.

That way, you determine the number of shares and the price per share at which you are willing to trade. Your order will appear in the order book among the buy or sell orders and is visible to the other participants. The other participants can then decide to accept your offer to buy or sell. They may fulfill your entire order or only part of it.

- Alternatively, you can accept buy or sell orders of other participants.

That way, you can select the best offer to buy or sell in the order book. You may decide fulfill the entire order or only parts of it.

Note: Buy and sell orders from the order book will not be matched automatically, even though the price might fit. To trade, you or another party always has to accept an outstanding offer.

Note: You also have the **possibility to sell shares short**. Short-selling means selling more shares than you have in your inventory. Example: Another subject offers to buy 5 shares, but you only have 3. With short-selling, you can still sell 5 shares. Your inventory is then -2 shares. You can buy back shares over the periods to return to a positive inventory. If you hold a negative inventory at the end of a period 12, then the following amount will be deducted from your cash balance: $(\text{number of shares}) \times (\text{the firm's retained earnings})$. Below you see a screen of the trading platform.

Periode 1 von 12 Verbleibende Zeit (Sekunden): 27

Transaktionspreise

180.00
140.00

Kaufgebote	
Volumen	Preis
2	110.00

Verkaufgebote	
Preis	Volumen
180.00	1
200.00	2

Gebotsabgabe

Geld: 2320.00
 Aktien: 5
 Letzter Preis: 140.00

Preis:
 Volumen:

Verkauf Stornieren Kauf Stornieren Kaufgebot Verkaufgebot

Typ	Preis	Volumen	Aktion
Gehandelt	140.00	1	Verkauf
Gehandelt	180.00	1	Verkauf

Alle Gebote

Gehandelte Gebote

Stornierte Gebote

Gelöschte Gebote

Sie können dafür stimmen, die Periode frühzeitig zu beenden. Sie können trotzdem weiter handeln, bis alle Ihre Stimme abgegeben haben, oder die Periode durch Zeitablauf endet. Bisherige Stimmen: 0/3

Für Periodenende stimmen

Summary of the sequence of events. Let us summarize the sequence of events in a market again:

1. You predict the average trading price of the upcoming trading period.
2. You can trade shares.
3. The firm announces the earnings of the period that just ended.
4. You receive an overview of your current wealth.
5. Trading starts all over again. After period 12, the market ends.

Next, we will play 3 practice rounds to familiarize you with the trading platform. The values in the practice rounds are *arbitrarily chosen* and have nothing to do with the actual markets that appear later on. The practice rounds will not count towards your final profits either. If you have any questions during the practice rounds, please do not hesitate to ask. After the practice rounds, we will conduct two quizzes and play a number of lotteries. You will receive 1 Euro per quiz that you answer correctly. At the end of the experiment, one of the lotteries will be selected and played out at random. Finally, after the quizzes, the first actual market starts, followed by the second market.

Summary of profit calculation. You have three different profit sources:

1. Your trading performance:

At the beginning of a market you receive an initial endowment of cash and some shares. At the end of the market, you might have some cash and hold some shares. Your final wealth in Taler is then determined as follows:

$$\text{Final Wealth} = \text{Cash} + \text{Number of shares} \times \text{Accumulated Earnings per Share}$$

Your wealth in Taler is then converted into Euros at the rate of 500:1. At the end of experiment, one of the markets will be chosen at random for payment.

2. Your forecast accuracy:

You receive 0.20 Euros for each forecast that is within a 10% range of the actual price. Your forecast accuracy from one of the markets will be selected at random for payment.

3. The lotteries and your answers to the quizzes.

Example:

Suppose your final wealth at the end of period 12 consists of 10 shares and 7.000 Taler Cash in one of the markets. The firm's accumulated earnings per share are 400 Taler. Your final wealth therefore amounts to: $7.000\text{Taler} + 10 \times 400\text{Taler} = 11.000\text{Taler}$. Your trading performance translates into $11.000\text{Taler} \times \frac{1}{500} = 22\text{Euros}$

Next, you have predicted the actual trading price correctly 5 times: $5 \times 0.20\text{Euros} = 1\text{Euro}$ Finally, you might have answered two quizzes correctly (2 Euros) and receive 5,60 from the lotteries.

Your final profit therefore amounts to: $22 + 2 + 5.60 + 1 = 30.60$ Euros