European University Institute
DEPARTMENT OF ECONOMICS

Essays in Empirical Economics

Rasmus Pank Roulund

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 20 May 2019

European University Institute
Department of Economics

Essays in Empirical Economics

Rasmus Pank Roulund

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Jérôme Adda, Bocconi University, Supervisor
Prof. Piero Gottardi, University of Essex
Prof. Rosemarie Nagel, Universitat Pompeu Fabra
Prof. Glenn W. Harrison, Georgia State University

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I Rasmus Pank Roulund certify that I am the author of the work Essays in empirical economics I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

Statement of inclusion of previous work (if applicable):

I confirm that chapter 1 was jointly co-authored with Nicolás Aragón and I contributed 50% of the work.

I confirm that chapter 3 was jointly co-authored with Nicolás Aragón in the revised edition of the thisis and I contributed the majority of the work (80%).

Signature and Date:
Rasmus Pank Roulund, May 5, 2019

European University Institute
Department of Economics

Essays in empirical economics

Rasmus Pank Roulund

February 23, 2018

Revised on:
March 19, 2019

Abstract

This first chapter is co-authored with Nicolás Aragón and examines how participant and market *confidence* affect the outcomes in an experimental asset market where the fundamental value is known by all participants. Such a market should, in theory, clear at the expected value in each period. However, the literature has shown that bubbles often occur in these markets. We measure the confidence of each participant by asking them to forecast the one-period-ahead price as a discrete probability mass distribution. We find that confidence not only affects price-formation in markets, but is important in explaining the dynamics of bubbles. Moreover, as traders' confidence grows, they become increasingly more optimistic, thus increasing the likelihood of price bubbles.

The second chapter also deals with expectations and uncertainty, but from a different angle. It asks how increased uncertainty affects economic demand in a particular sector, using a discrete-choice demand framework. To investigate this issue I examine empirically to what extent varying uncertainty affects the consumer demand for flight traffic using US micro demand data. I find that the elasticity of uncertainty on demand is economically and statistically significant.

The third chapter presents a more practical side to the issue examined in the first chapter. It describes how to elicit participants' expectations in an economic experiment. The methodology is based on Harrison et al. (2017). The tool makes it easier for participants in economic experiments to forecast the movements of a key variable as discrete values using a discrete probability mass distribution that can be "drawn" on a virtual canvas using the mouse. The module I wrote is general enough that it can be included in other economic experiments.

Contents

I would like to thank (in no particular order) Andrea Mattozzi, Benedikt Kolb, Christian Michael, David K. Levine, Elizabeth Tepper, Filip Lazaric, Francisco Roldan, Gabriele Camera, Gabriele Facchini, Giovanni Giusti, Glenn W. Harrison, Hrvoje Stojic Jerome Adda, John Duffy, Jonathan Chapman, Marcelo Fernandez, Matthias Sutter, Nicolás Aragón, Pablo Forero, Piero Gottardi, Robert Goodhead, Rosemarie Nagel, William N. Evans, and Yaron Lahavragón for helpful comments and support.

Chapter 1.

Certainty and Decision-Making in Experimental Asset Markets

Co-authored with: Nicolás Aragón (Universidad Carlos III de Madrid)

Abstract

This paper examines how traders' certainty and market certainty affect outcomes in an experimental asset market with known fundamental values. In this type of market, prices usually present large deviations from the fundamental value; in other words, bubbles are known to occur. We measure beliefs by asking participants to forecast the one-period-ahead price as a discrete probability mass distribution. We define certainty as the inverse of the dispersion of beliefs for each trader, and also create a market-wide measure of this to measure agreement across traders. We find that certainty affects price-formation and is also important in explaining the dynamics and size of the bubble. Moreover, as traders are successful they become increasingly more certain in their beliefs, even if these are on non-fundamental values, thus increasing the likelihood of price bubbles.

Expectations are of central importance to understanding price formation and fluctuations in asset markets (see, for example, Fisher and Statman, 2000). In particular, heterogeneity of expectations is generally necessary to break the no-trade theorem (Tirole, 1982a).

Trader expectations have largely been studied in experimental setups. Measuring them in practice is challenging, as fundamental values are unknown. A major shortcoming of the experimental literature is that it tends to focus on eliciting point forecasts.[1] Because of this, the role of certainty of beliefs cannot be examined.

In this paper we propose a setup that enables us to understand the role of belief heterogeneity

1 Notable exceptions are Deaves et al. (2008), Biais et al. (2005), Michailova and Schmidt (2016), and Kirchler and Maciejovsky (2002), who use confidence bands, as discussed in Section 1.

and trader certainty in the formation and fluctuation of prices in experimental asset markets. We do this by eliciting traders' distributions of beliefs.

Our main contribution is to examine the effect of both traders' dispersions of beliefs and the market dispersion of beliefs on price dynamics in an experimental asset market. We elicit a distribution of beliefs in future prices, and define trader certainty as the inverse of the dispersion of those beliefs. This allows us to assess how tightly held the expectations about future prices are. We also create a market-wide measure of market certainty, which allows us to explore trader agreement.[2] The main results are that trader certainty matters for predicting price movements, and that market certainty is associated with larger bubbles.

We build on the workhorse of experimental finance (following Smith, Suchanek and Williams, 1988, SSW henceforth). In this setup, prices are known to deviate largely from fundamental values; i.e., bubbles are typically observed. We use the extension by Haruvy et al. (2007) and elicit beliefs based on Harrison et al. (2015) and Harrison et al. (2017). Subjects participate in an experimental session consisting of two or three markets, each of which lasts for either 15 periods or 12 periods, respectively. There is a single asset in the market, which pays a random dividend from a known, fixed distribution. Trading is conducted as follows: in each period, agents can buy and sell assets. Prices are determined via a call market and a unique price clears the market. In addition, before trade is conducted, traders are asked to forecast the price. Traders are endowed with 20 tokens, which they can allocate to different price ranges to assess the likelihood of each range. In this way, we can elicit distributions of beliefs about prices, which allows us to analyze the impact of trader certainty on market dynamics. We also aggregate certainty at the market level, creating a measure of market certainty and (dis)agreement.

Since agents have two tasks that are incentivized, they may hedge between them. This can result in distortions in reported beliefs, as shown by Rutström and Wilcox (2009).[3] Following the literature, we avoid any potential hedging between the two tasks by rewarding participants randomly for only one of them at the end of the game.

We first analyze how certainty evolves at the individual level and show that traders who are successful in predicting become more certain in their predictions. Secondly, we find that traders' expectations converge in mean as they are more experienced, but belief heterogeneity persists. Thirdly, we find that prices become more disconnected from fundamental values when traders' beliefs are less scattered (i.e., the size of the bubble is amplified). Finally, we find that certainty in beliefs has predictive power over future price movements.

2 A proper definition of confidence (and overconfidence) presents several theoretical and empirical challenges, which we review in the literature review. Given the complexity of the setup, we cannot fully disentangle these notions, so our focus is on the trader dispersion of beliefs or its certainty.

3 There are other papers that show the effect of hedging. See Armantier and Treich (2013a) and Blanco et al. (2010). Hanaki et al. (2018) show that there is indeed different behavior when subjects participate in asset markets and need to both trade and make forecasts.

The rest of the paper is structured as follows. Section 1.1 reviews the theoretical and empirical literature, with particular emphasis on bubbles in asset markets. Section 1.2 describes the hypotheses that guide our design, presented in Section 1.3. Section 1.4 provides an overview of the data. We present the results in Section 1.5. Section 1.6 concludes.

1.1. Literature Review

The experimental literature on bubbles dates back to the presentation of the SSW framework. In this setup, typically nine participants are endowed with different amounts of cash and assets, and can trade during several 15-round markets. The SSW setup has been the workhorse for experimental asset market research for the last 30 years, covering many areas of trading, including the role of traders' characteristics (experience, education, sentiments, etc.), public announcements, liquidity, short selling, dividends, capital gains taxes, and insider information, among others (see Palan, 2013, and Powell and Shestakova, 2016, for recent surveys).[4] A common feature of these studies is that bubbles emerge even though the fundamental value of the asset is known in every round.

Haruvy et al. (2007) address the impact of long-run expectations in that setup. Other papers have designed experiments exclusively to isolate expectations (see Heemeijer et al., 2009; Hommes et al., 2008). We build upon Haruvy et al. (2007), given the large body of literature that the SSW setup has inspired, making our results more broadly comparable. Within that setting, we elicit the distribution of beliefs instead of point forecasts. We contribute to this line of research by focusing on two dimensions: the dispersion of beliefs across traders and the certainty of beliefs for individual traders. We also focus on heterogeneity across traders and individual trader certainty.

Individual Trader Certainty. Our paper is the first to elicit a whole distribution of beliefs for each trader in an SSW setting. Previously, papers had in general elicited expectations as point forecasts with a confidence band (as in Haruvy et al., 2007, who focus on long-run expectations). These papers contributed to the literature on overconfidence.

From a theoretical standpoint, overconfidence has been used to explain the occurrence of bubbles (see, for example, Scheinkman and Xiong, 2003). Agents believe that their information is superior to that of the general market, and fail to adjust their expectations as they observe others' beliefs. However, overconfidence leads to behavior that is not necessarily distinguishable from behavior stemming from rational information processing via Bayesian updating (Benoît and Dubra, 2011).

4 Other types of experimental asset markets have also been employed in the literature to study the occurrence of crashes, and topics such as information dispersion, market inefficiency, and the impact of future markets, among others. See Sunder (1995) and Noussair and Tucker (2013) for surveys on asset market experiments.

Moore and Healy (2008) discuss three different theoretical forms of overconfidence. For our purposes, the relevant form of overconfidence would be miscalibration, which is the excess accuracy of one's beliefs, in this case with respect to future prices. This implies that people choose an overly narrow confidence interval when asked for a range that contains the true value (Alpert and Raiffa, 1982). Several papers have attempted to measure this in experimental asset markets. Biais et al. (2005) use the Plott and Sunder (1988) framework of asset trading with noisy private signals, where there is a prevalent winner's curse. In this setup, they elicit confidence as an interval. They analyze overconfidence as miscalibration to explain why subjects fail to take into account the winner's curse risk. When conditional uncertainty about the value of the asset is high, rational agents will recognize this. By contrast, miscalibrated traders will be less aware of this, and thus show excessive confidence in their assessment of the value of the asset. Michailova and Schmidt (2016) and Kirchler and Maciejovsky (2002) investigate the effect of overconfidence (as miscalibration) and risk aversion in an SSW framework, with no informational asymmetries. In both papers, subjects received a pre-experimental overconfidence score and were assigned two treatments based on the score. After the experiment, subjects participated in a risk aversion measurement test. The papers find that there is a significant effect of overconfidence on individual outcomes. Deaves et al. (2008) also analyze a trading market with private signals, where they attempt to disentangle the forms of overconfidence and focus on gender differences in behavior.

A shortcoming of the papers measuring overconfidence in financial markets is that there are many variables that are intertwined. Importantly, it is not obvious how to disentangle the role of learning and the role of higher order expectations. In other words, the setups are too complex to elicit overconfidence properly. This is the reason why we prefer to focus on the dispersion of beliefs or trader certainty.[5] Our chosen method for eliciting expectations allows participants to report different intensities of beliefs, or disjoint distributions, and not only confidence bands. This is particularly relevant in asset markets with bubbles, as prices can follow either a fundamental or a non-fundamental trend. Moreover, our focus is on the predictive power of certainty and its dynamics.

Dispersion of beliefs across traders. This topic has received theoretical, empirical[6] and experimental attention.

The theoretical strand focusing on belief heterogeneity breaks the no-trade theorems, since it argues that private information alone cannot explain bubbles (Milgrom and Stokey, 1982). Moreover, Tirole (1982b) shows that speculative bubbles—those in which speculative investors buy an overpriced asset to sell it to a greater fool before the crash—cannot arise in rational expectations models. Thus, many papers have relied on so-called "irrational" behavior or myopia to explain them (see, for example, Abreu and Brunnermeier, 2003). Most of the models that explain bubbles focus on two issues: coordination and dispersion of opinion. If a trader

5 In our case, the role of learning should be, on average, the same across sessions, given that all participants were recruited under the condition of being new to this type of game.

6 See for instance, Hommes et al. (2017)

expects the asset price to soar, she may be willing to buy the asset even at a price above its fundamental value, given the possibility of reselling at a higher price at a later time. As long as the moment of the crash is unknown, traders might be tempted to ride the bubble, hoping to sell before the bubble bursts. Central to this line of research is the dispersion of beliefs. In particular, Varian (1985) and Miller (1977) highlight the effect of dispersion of beliefs on trading volumes. Brock and Hommes (1997) provide a model of asset trading based on the interaction of traders with heterogeneous expectations, giving rise to dynamics akin to those observed in real-world stock prices. Boswijk et al. (2007) and Barberis et al. (1998) focus on the psychological factors at play in asset trading, specifically overconfidence and framing effects. From a psychological point of view, Minsky (1992) has argued that when investors agree on forecasts, optimism drives up prices. In other words, agreement in expectations leads to more severe bubbles. Our experimental design allows us to analyze the impact of conformity among traders and asset price dynamics.

Hommes (2011) surveys laboratory experiments that deal with the heterogeneity of beliefs and analyzes their data. The main finding is that heterogeneous expectations are necessary to explain aggregate outcomes. Boswijk et al. (2007) analyze heterogeneous expectations in experimental markets designed to have positive and negative feedback loops. In our paper, we examine not only the heterogeneity across agents, but also the certainty of each agent (i.e., how tightly held the beliefs are). Importantly, our mechanism of belief elicitation allows us to capture the intensity of the disagreement across traders. Finally, our SSW setup does not have any built-in mechanism of positive or negative feedback loops.

Baghestanian et al. (2015) develop a model of heterogeneous traders. Using experimental data, they identify subjects as noise traders (who buy or sell, on average, at the previous price), fundamental traders (who buy when the price is below the fundamental value and sell above the fundamental value) and speculators (who buy to re-sell to noise traders above the fundamental value). They emphasize individual modeling to capture aggregate features of the data, and show that these strategies generate bubble-crash patterns. In this paper, we emphasize heterogeneous beliefs while remaining agnostic about the typology of traders; and we show how these beliefs reinforce each other and explain asset price dynamics.

Carlé et al. (2019) use the data from Haruvy et al. (2007) to analyze heterogeneous expectations in a setup where agents have the same information. They find that trade happens because more optimistic traders purchase from less optimistic ones. Importantly, agents have heterogeneous beliefs and act accordingly. Their main finding is that divergence of opinion has an effect on mispricing. Our results are aligned with this finding, but our setup allows us to examine not only the effect of the dispersion of beliefs (and its intensity) across traders, but also the effect of trader certainty.

1.2. Hypotheses

The following hypotheses guide the experimental design and our analysis.

Our first hypothesis concerns the dynamics of individual certainty, which we define as the (inverse of the) dispersion of beliefs at the individual level. We expect that, as traders successfully predict the sign and the magnitude of the price movement from one period to the next, they will have more concentrated beliefs going forward. This hypothesis is consistent with the theoretical literature on trader behavior in asset markets (see, for example, Scheinkman and Xiong, 2003; Harrison and Kreps, 1978), as well as with Bayesian updating.

Hypothesis 1. *Traders who predict the correct sign of the price movement will be more certain about their predictions in the following period.*

A well-known result is that bubbles disappear as traders participate repeatedly in different markets. This implies that, eventually, equilibrium prices closely follow the fundamental value. A typical interpretation is that markets converge to the full-information rational expectations equilibrium (Palan, 2013).[7]

A potential explanation for the convergence of empirical prices to the fundamental value is that agents' expectations align over time. Inspired by these facts, we expect traders' expectations to converge. However, even though expectations may converge, the fact that trade normally persists would hint at a violation of the no-trade theorem. Thus, we expect beliefs to converge in mean, but belief heterogeneity to remain even in the later periods.

Hypothesis 2. *Traders' expectations converge in mean to the fundamental value as traders participate in more markets, but the heterogeneity of beliefs remains.*

One of the main objectives of this paper is to analyze whether the agreement in expectations affects the size of the bubble. The Financial Instability Hypothesis (Minsky, 1992) states that that success in financial markets leads to reckless behavior. Moreover, Minsky (1992) argues that markets tend to aggregate information in such a way that the optimism of a given trader reinforces the optimism of other traders.

In other words, if all agents believe the price will be high, this will push the price up. However, if beliefs are scattered, the more optimistic traders will become more pessimistic, and vice versa. In our setup, we shed some light on this hypothesis by studying whether the dispersion of beliefs at the market level and trader certainty are related to an increased size of bubble.

7 A notable exception is Hussam et al. (2008). Here, the authors show that bubbles can be rekindled by changing the fundamental parameters characterizing the market. They argue that this corresponds to (unexpected) technological changes.

Hypothesis 3. *Increased initial agreement in traders' expectations, measured as the overlap in traders' expectations, leads to increased bubble sizes* ceteris paribus*, and more initially certain traders will increase bubble size.*

Our last hypothesis relates to the predictive power of market confidence. It is well known that expectations have predictive power for market outcomes. In an experimental setup, Haruvy et al. (2007) have found that the mean of traders' expectations partly explains price movements. We are interested in whether the dispersion of beliefs, as a proxy for market confidence, can help predict future price movements. In other words, we want to address whether higher moments of the distribution of expectations matter to asset price dynamics.

Hypothesis 4. *Information on traders' certainty has predictive power on future price movements.*

In the next section we present the experimental design chosen to address our hypotheses.

1.3. Experimental Design

Our experimental design follows Haruvy et al. (2007), though it differs in two crucial aspects: (i) we elicit certainty in forecasts; (ii) the payment is based on either forecast performance *or* trade performance. We discuss both changes below.

In each session 12 traders are grouped together (Haruvy et al., 2007, use 9 traders). They participate in two or three sequential *markets*, each consisting of 15 or 12 trading periods, respectively. Each market has two stock valuables: an asset, denominated in *shares*, and money, denominated in *points*. Within a market, the holdings of shares and points are transferred from one period to the next. Each trader is given an endowment of shares and money at the beginning of each market.[8] In each period, the asset pays an IID random dividend in points from a known and fixed distribution. The timing of a given period t is summarized in Table 1.1. Participants have two tasks: trade and predict. Traders can buy and sell shares in each period. Based on the orders, an equilibrium price is computed. In addition, at the beginning of each period, subjects are requested to forecast the equilibrium price. We elicit approximations of the traders' distributions of beliefs as described in section 1.3.2. Traders are rewarded according to either the accuracy of their forecast or according to how much money they have accumulated. The reward is determined individually at the end of each market and transformed into euros at a rate of 85 points per euro.

8 Denote an initial endowment of f Points and a shares of the assets by (f, a). Then four participants are endowed with $(112, 3)$, other four participants are endowed with $(292, 2)$, and the remaining participants are endowed with $(472, 1)$. Thus, a total of 24 shares of the asset and 3504 Points in the first round, and up to 14304 Points in the last round.

Table 1.1.: Summary of the timing of round t, where T is the total number of rounds.

Start of period t	Traders predict the equilibrium prices.
	Traders make buy and sell orders.
	The equilibrium price is calculated and trade is conducted.
	The asset dividend is drawn and added to cash holdings.
End of period t	Forecast earnings are computed based on accuracy of forecasts.

The experiment was conducted at the Behavioral Sciences Laboratory at Universitat Pompeu Fabra, Barcelona (BESLab) and utilized software developed by Chen et al. (2016) and Bostock et al. (2011). Subjects were typically bachelor students from a wide range of disciplines with no prior experience with experimental asset markets. Each session lasted for approximately 2.5 hours, of which 45 minutes was instructions. [9] Before being able to start the experiment, each trader had to complete an interactive test that demonstrated understanding of the basic mechanisms of the experiment. The average earnings per trader were approximately €25.

Below we describe the setup in greater detail.

1.3.1. The asset market

The asset pays a dividend at the end of each period for each share. The dividend is drawn from the set $\{0, 4, 14, 30\}$ with equal probability. It is common knowledge that the expected dividend is 12 points. Likewise, it is common knowledge that the "fundamental value" of a share of the asset in the beginning of period $t \in \{1, \dots, 15\}$ is $f(t) = 12\,(15 - t + 1)$.

In each period, traders can buy and sell the asset, subject to the "no borrowing" constraint via a call market. As in Haruvy et al. (2007) the price is determined by call market rules to ease the task of predicting future prices, as the more commonly employed double auction rules can lead to multiple prices. In each period, each trader submits a buy order and a sell order. An order consists of a price p and a quantity x. A buy order specifies the maximum quantity x_b of shares that the trader is willing to purchase at a price lower or equal to p_b. Likewise, a sell order specifies the maximum quantity x_s of shares that the participant will offer if the price is at least p_s. All bids and asks are aggregated into demand and supply curves, respectively, and the equilibrium price p^* is determined.

The equilibrium price algorithm is designed to maximize trade volume, given buy and sell bids. If there is no intersection between the supply and demand curve, the equilibrium price

9 The instructions are available in English in Appendix A.2, and in Spanish at `http://pank.eu/experiment/introduction.html`

is set to the one plus the highest offer, i.e. $p^* = 1 + \max_i \{p_{si}\}$.[10] If there is a unique volume-maximizing price, then this is the equilibrium price. Otherwise, we check for prices that make the difference between supply and demand zero within the set of volume-maximizing prices. If no such price exists we check whether all prices lead to excess supply or demand and take the maximum or minimum price, respectively. Finally, if the price has not been found in any other way, we set it to the maximum price that minimizes the difference between supply and demand.

Participants who submitted buy order with $p_b \geq p^*$ will purchase assets from participants with sale offers satisfying $p_s < p^*$. If demand exceeds supply at the margin, realized trades at the margin are randomized.

1.3.2. Eliciting traders' beliefs

The main methodological contribution of this paper is that we elicit traders' distributions of beliefs using a elicitation tool similar to the one proposed by Harrison et al. (2017) in an asset market. The traders participating in the Haruvy et al. (2007) experiment are not able to express the degree of certainty in their forecasts or put weight on two disjoint distributions. The latter may be particularly relevant when forecasting the burst of a bubble; some weight may be allocated to the bubble price path and some weight to the fundamental price path. In our experiment participants can express a good approximation of their entire belief structure.

Our elicitation tool is similar to the tool found in Harrison et al. (2017). The tool is displayed in Figure 1.1. Participants are shown a grid with prices on the primary axis and percentages along the secondary axis, and are asked to predict the price in next period. To do so, they are endowed with 20 tokens, each representing 5% certainty. Prices are binned in intervals of 10 along the primary axis. Along the secondary axis participants are shown bins of 5% intervals. By clicking on a particular grid point, subjects allocate their tokens. Participants are not allowed to continue unless they assign all their 20 tokens. Once the tokens have been allocated the forecast can be "finalized" and the scores are shown. Subjects can revise the forecast.[11]

Participants are then rewarded in accordance to the accuracy of their forecasts. However, ensuring that expectations are correctly elicited is a difficult task. We will first describe the elicitation process and then discuss how we circumvent the main issues.

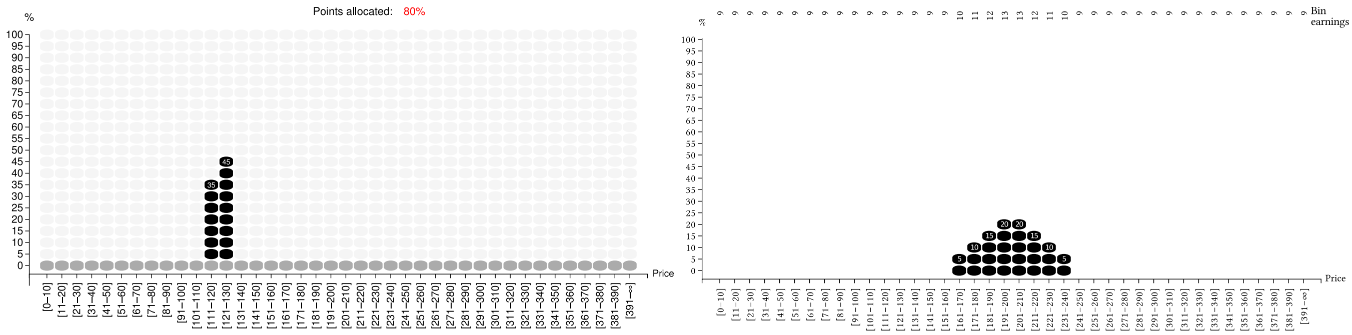
The price range is partitioned in $k = 1, \dots, K$ bins. The maximum price was set such that the

10 This was the approach taken by Haruvy et al. (2007).

11 One difference from the tool used in Harrison et al. (2017) is that it displays scores as tokens are being distributed.

Figure 1.1.: An example of a constructed forecast of prices in a given period.

(a) Before the forecast is finalized. (b) A finalized forecast.



maximum price observed in the literature was not binding. An example of a constructed probability mass function is shown in Figure 1.1.[12]

We denote the beliefs (tokens) in a particular price interval k for a particular period τ as $s_{\tau k}$. Participants are asked to choose beliefs for every remaining period in each one of the k bins, thus assigning probabilities to each future potential price. A full report consists of belief allocations for each interval $s_{\tau k} = (s_{\tau 1}, s_{\tau 2}, \dots, s_{\tau K})_{\tau=t}^{j=T}$. In this way we can elicit certainty from the forecast as in Harrison et al. (2017). The setup requires participants to assign zero probability to some bins[13] and allows for arbitrary probability distributions of forecasts, which may be disjoint. This is important because participants may want to bet on two different equilibrium prices paths. To see this, consider the case where the forecaster expects that a bubble continues for one more period with probability q and that the bubble bursts and the market crashes with probability of $1 - q$. This gives rise to a bimodal prediction of the future price given the fact that some subjects may put some weight on the fundamental value and some weight on the bubble value.

The difference between our baseline treatment and Haruvy et al. (2007) is that we explicitly elicit a measure of the dispersion of beliefs. To incentivize careful predictions of the certainty levels we make payoff of the prediction contingent on the accuracy of forecasts. Scoring rules are procedures that convert a "report" into a lottery over the outcomes of some event. These rules are a way of translating reported beliefs into earnings based on the actual outcome. Payoff from forecasting follows the Quadratic Scoring Rule (QSR),

$$S = \kappa\Big[\alpha + \beta\Big(2 \times r_k - \sum_{i=1}^{K} r_i^2\Big)\Big], \tag{1.1}$$

12 An example of the actual software implementation used in the experiment may be found at `http://pank.eu/experimenth/hist2.html`.

13 The price grid goes from zero to 400 points and has a ten-point span for each grid point. As such, participants must consider where in this range prices will fall even if using a diffused distribution.

following (Matheson and Winkler, 1976). This reward score doubles the report allocated to the true interval and penalizes depending on how these reports are spread across the K intervals. α, β and κ are calibrated constants. In particular, participants are rewarded firstly according to whether the realized price falls within the correct band and secondly according to the precision of this band. The main difference between (1.1) and the Haruvy et al. (2007) setup, is that here the confidence band is determined by the participant. This scoring rule has several desirable properties which are discussed in the next subsection.

1.3.3. Risk, Ambiguity and Hedging

There are two main issues to account for when eliciting beliefs. The first is the possibility of hedging. The second is the interconnection in reported beliefs between ambiguity, risk aversion and the dispersion of beliefs.

Hedging refers to the fact that, for example, subjects may trade at optimistic values and forecast at pessimistic values in order to ensure a less variable payoff.

This may create distortions in reported distributions, as shown by Armantier and Treich (2013a), Blanco et al. (2010) and Rutström and Wilcox (2009). We follow the literature and randomize the payoff given to the agent at the end of each market.[14] We avoid any potential hedging between the two tasks by rewarding participants randomly at the end of the game. In particular, participants are randomly rewarded either according to the accuracy of their predictions, as in Haruvy et al. (2007), *or* according to their earnings in the asset market. In the initial pilot sessions payoffs were calibrated to match the same expected earnings in both types of task. In this way, we are able to elicit more detailed information about beliefs and address the impact on price dynamics.

The risk attitudes may impact the incentive to report one's subjective probability truthfully in equation (1.1). Harrison et al. (2017) characterize the properties of the Quadratic Scoring Rule when a risk-averse agent needs to report subjective distributions over continuous events. For empirically plausible levels of risk aversion it is possible to elicit reliably the most important features of the latent beliefs without calibrating for risk attitudes. The qualitative effect of greater risk aversion, when eliciting continuous distributions, is to cause the reported distributions to be "flatter" than the true distributions. In particular, the Scoring Rule has the following properties:

1. The individual never reports a positive probability for an event that does not have a positive subjective probability, independently of the risk attitude;

14 Another possibility is doing "small" payments in order not to distort the incentives. However, since it is the main focus of our paper we decided to incentivizing this activity properly.

Table 1.2.: Overview of the sessions, number of markets and number of participants.

Session	Markets	Rounds	Participants	Assets
A	2	15	12	27
B	2	15	12	27
C	3	12	12	27
D	3	12	12	27
E	2	15	12	27

2. Events with the same subjective probability have the same reported probability;

3. The more risk-averse the agent is, the more the reported distribution will resemble a uniform distribution over the support of the true latent distribution.

It must be mentioned that it is also possible to use a binary lottery procedure, developed by Harrison et al. (2015) to risk-neutralize the participants.[15] This would be particularly interesting due to the fact that the above-mentioned procedure does not necessarily work with rank-dependent-utility participants, as what is recovered are the weights and not necessarily the probabilities. This would have added another layer of complexity to an already complex experiment. With all these caveats, we assume away risk preferences. However, given that the effect of risk aversion is to report a distorted distribution on the true range of probabilities, we can certainly infer a lower bound on the dispersion of beliefs.

1.4. Overview of experimental data

In this section we show the basic properties of the data from our study conducted at the BESlab, Universitat Pompeu Fabra, Barcelona. A total of five sessions were conducted.[16] The summary of number of participants and markets is shown in Table 1.2. Sessions A, B and E had two markets, each with 15 rounds. Sessions C and D had three markets each consisting of 12 rounds. All sessions had 12 participants. Likewise, all sessions had a total of 27 tradable assets.

15 Moreover, Harrison and Ulm (2015) develop a framework to recover the parameters from the utility function for very general families of functions.

16 A couple of pilot sessions were also conducted. The data from these sessions is not used, as they either did not conclude, were conducted with fewer participants, or were conducted using a vastly different version of the experimental software.

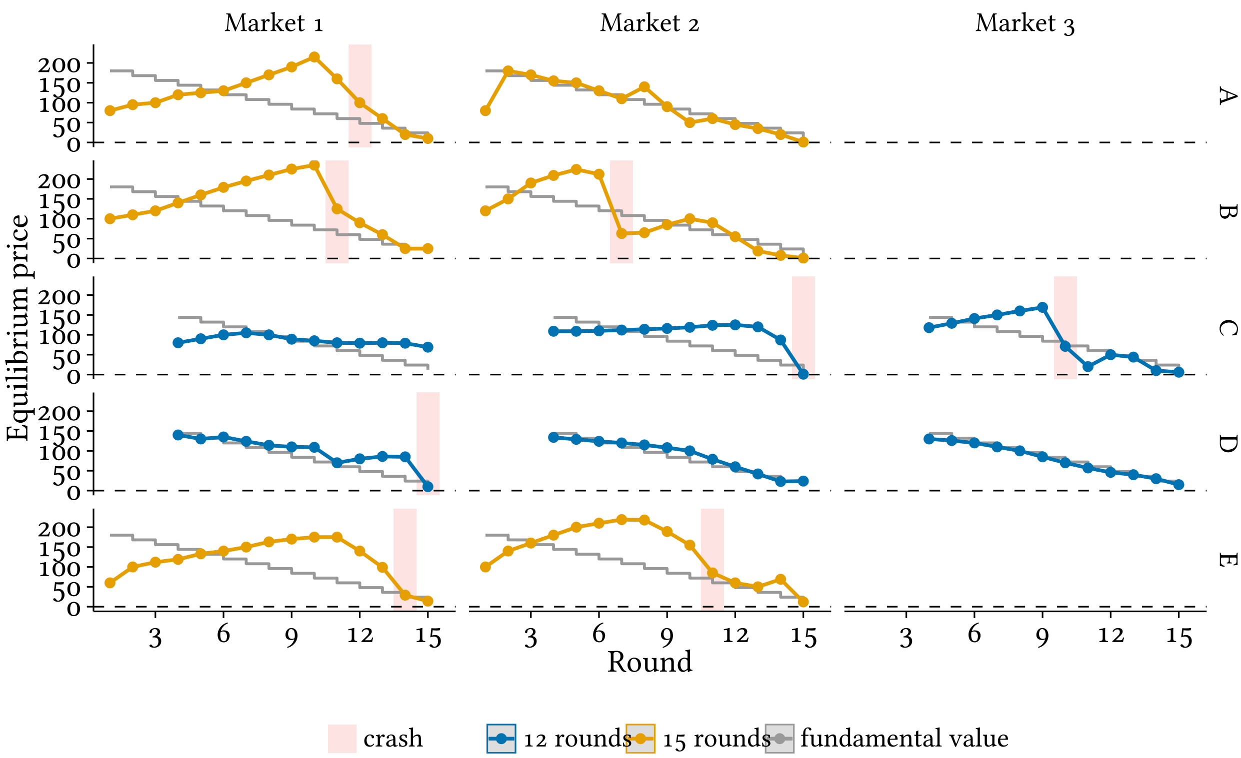


Figure 1.3.: The development of equilibrium prices over markets and market periods. The gray step-lines display the fundamental value of the asset.

1.4.1. Summary of the trade data

We first turn to the descriptive statistics of the main variables in the experiment. As argued above, traders react and set the price in this experiment. The period-to-period development of the price for each of the sessions is shown in Figure 1.3. The lines and the dots in Figure 1.3 show the price movement. Curves and dots are color-coded depending on the total number of rounds per market. The gray step lines display the fundamental value of an issue of the asset in a given round. Note that sessions with 12 rounds per market have been shifted to start at period 3 to align the final period across graphs.

In a market with perfect foresight and fully rational players the price curve should coincide with the fundamental value curve. In addition, the graph shows periods where the price "crashes", corresponding to a fall in price of more than 60 points. The exact prices may be found in Table 1.3b, where the fundamental value is also shown. The price movements display similar trends to those found in similar experiments, such as Haruvy et al. (2007). The peak price typically occurs around period 10 to 11 in the sessions with 15 rounds. As can be seen from the graph, the markets clearly exhibit bubble tendencies, as is found in most similar experiments. In our second market bubbles still occur in session 2, though less so than in the first market. This indicates a similar learning process to the one described in Haruvy et al. (2007), so we

Table 1.3.: Realized price in each round per market and per session. f_t denotes the fundamental value for period t. Note that we have shifted down sessions with 12 rounds, as the fundamental value depends on the remaining number of rounds.

		Market 1					Market 2					Market 3	
t	f_t	A	B	C	D	E	A	B	C	D	E	C	D
1	180	80	100			60	80	120			100		
2	168	95	110			100	180	150			140		
3	156	100	120			112	170	190			160		
4	144	120	140	80	140	119	155	209	109	134	180	118	130
5	132	125	160	90	130	133	150	224	109	129	200	129	126
6	120	130	179	100	135	140	130	212	110	124	210	141	120
7	108	150	195	105	124	150	110	63	112	120	219	150	110
8	96	170	210	100	114	163	140	65	114	115	218	160	100
9	84	190	225	89	110	170	90	85	116	108	189	169	85
10	72	215	235	85	109	175	50	100	119	100	155	71	70
11	60	160	125	80	70	175	60	90	124	79	85	20	57
12	48	100	90	79	80	140	45	55	125	60	60	50	46
13	36	60	60	80	86	99	35	19	120	42	50	44	40
14	24	20	25	79	85	29	20	8	87	23	69	10	30
15	12	10	25	69	10	14	1	1	1	24	12	6	15

conclude that the proposed elicitation does not affect behavior.

There is a fair bit of variation between sessions. While the three sessions of 15 rounds, sessions A, B, and E, show a typical hump shape typical to this type of asset market experiments, the two sessions with 12 rounds, sessions C and D, show a less characteristic shape. Session D seems to almost converge to the fundamental value within the first market, and completely converges in the second market and the third market. Across sessions, the market crash tends to be in a later round in the first market, compared with market crashes in subsequent markets.

In each round, traders put forward a bid-quantity tuple and an ask-quantity tuple. Recall that bids are the maximum prices buyers are willing to buy at, while asks are the minimum prices sellers are willing to sell at. Based on this, an equilibrium price is computed in each round using the mechanism described in the previous sections and the realized prices shown in Table 1.3. In addition to the prices, the realized traded quantities are shown in Table 1.4. The number of traded assets is not shown to participants during the experiment. For sessions with 12 rounds per market the 12th and last round has been aligned with the last round in the sessions with 15 rounds. As the table shows, the trade volume is generally above zero, with a few exceptions in session C.

Table 1.4.: Realized trade per round, per market, and per session.

		Market 1					Market 2					Market 3	
t	f_t	A	B	C	D	E	A	B	C	D	E	C	D
1	180	80	100			60	80	120			100		
2	168	95	110			100	180	150			140		
3	156	100	120			112	170	190			160		
4	144	120	140	80	140	119	155	209	109	134	180	118	130
5	132	125	160	90	130	133	150	224	109	129	200	129	126
6	120	130	179	100	135	140	130	212	110	124	210	141	120
7	108	150	195	105	124	150	110	63	112	120	219	150	110
8	96	170	210	100	114	163	140	65	114	115	218	160	100
9	84	190	225	89	110	170	90	85	116	108	189	169	85
10	72	215	235	85	109	175	50	100	119	100	155	71	70
11	60	160	125	80	70	175	60	90	124	79	85	20	57
12	48	100	90	79	80	140	45	55	125	60	60	50	46
13	36	60	60	80	86	99	35	19	120	42	50	44	40
14	24	20	25	79	85	29	20	8	87	23	69	10	30
15	12	10	25	69	10	14	1	1	1	24	12	6	15

Table 1.5.: Summary statistics describing each market.

		1					2					3	
	All	A	B	C	D	E	A	B	C	D	E	C	D
Peak price	170.9	215.0	235.0	105.0	140.0	175.0	180.0	224.0	125.0	134.0	219.0	169.0	130.0
Peak period	5.6	10.0	10.0	4.0	1.0	11.0	2.0	5.0	9.0	1.0	7.0	6.0	1.0
Total trade	50.5	57.0	62.0	54.0	49.0	75.0	37.0	58.0	34.0	43.0	60.0	33.0	44.0
Turnover	1.9	2.1	2.3	2.0	1.8	2.8	1.4	2.1	1.3	1.6	2.2	1.2	1.6
Amplitude	2.3	2.5	2.7	5.2	2.7	2.6	1.4	1.7	3.5	1.1	2.3	1.7	0.3
Normalized deviation	65.2	92.8	114.0	59.5	36.0	149.6	20.3	82.6	51.4	20.8	111.4	36.6	7.7
Boom duration	7.8	8.0	11.0	8.0	9.0	11.0	8.0	4.0	8.0	8.0	12.0	4.0	3.0
Upward trend	4.7	9.0	9.0	3.0	2.0	9.0	1.0	4.0	7.0	1.0	6.0	5.0	0.0
Relative abs. deviation	0.4	0.6	0.6	0.4	0.3	0.6	0.2	0.4	0.5	0.2	0.6	0.3	0.1
Relative deviation	0.2	0.2	0.4	0.1	0.3	0.2	-0.0	0.1	0.3	0.1	0.4	0.1	-0.0
Avg. abs. forecast bias	2.7	2.8	2.9	2.2	2.6	2.7	3.3	3.2	2.0	2.4	2.1	3.5	2.3
Avg. forecast bias	1.6	2.0	1.5	2.1	2.4	2.6	0.7	1.4	0.8	2.4	0.2	1.4	2.3

The raw bid and ask data used to calculate the equilibrium price are shown below in Figure 1.4. The graph shows the demand and supply at each potential price as well as the realized price. A larger point indicates larger bids/asks at that point. Based on the pricing mechanism and the raw bids and asks, the equilibrium price is determined.

In Table 1.5 we display summary statistics of each market.

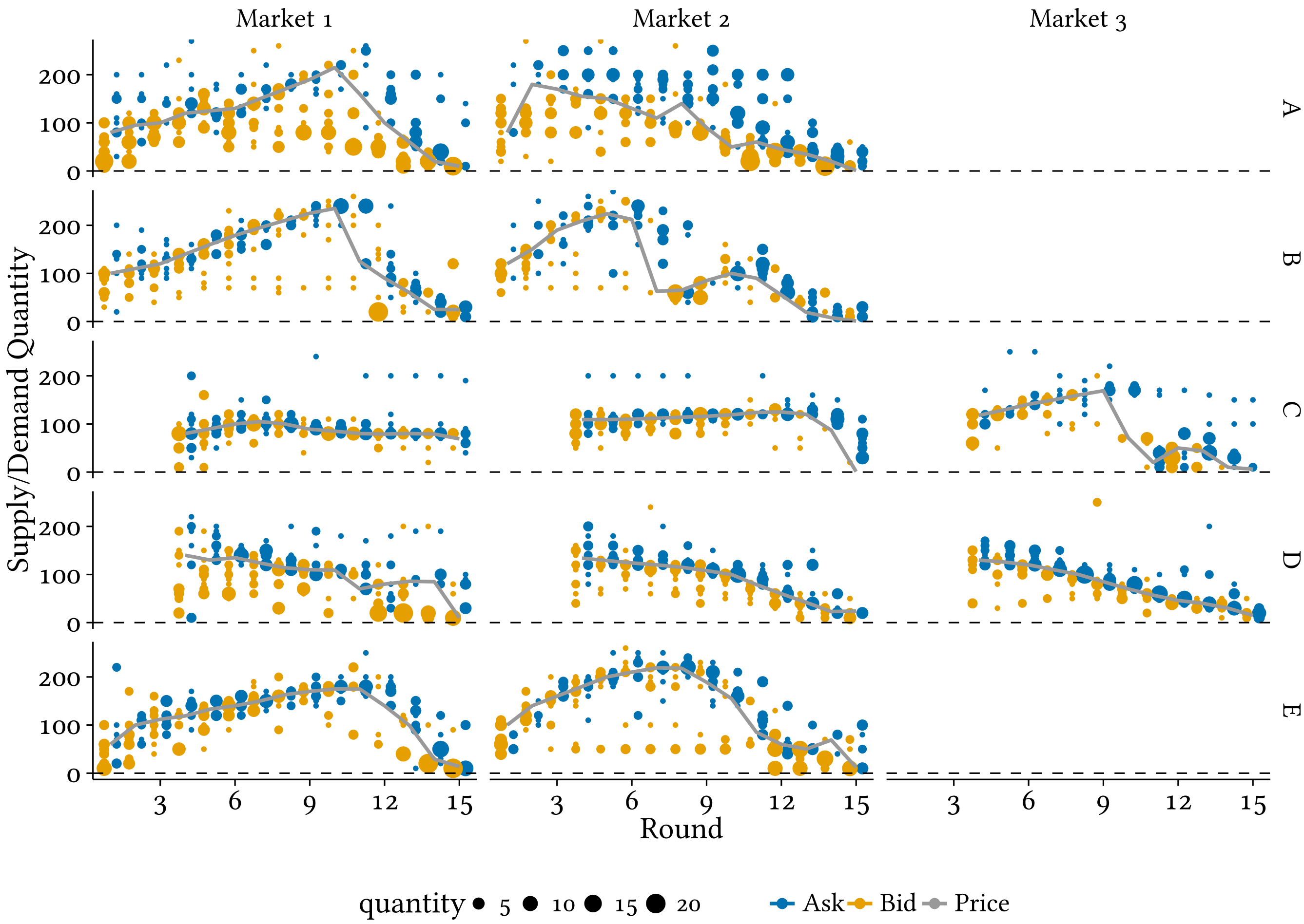


Figure 1.4.: Supply and demand per session, market and round.

1.4.2. Expectation data

The main difference in this experiment compared with the previous literature is the way forecasts are elicited. As discussed above, participants forecast the price development at the beginning of each round using the tool shown in Figure 1.1. Some examples of the forecasts that subjects produced are shown in Figure 1.5. The figure shows forecasts conducted by

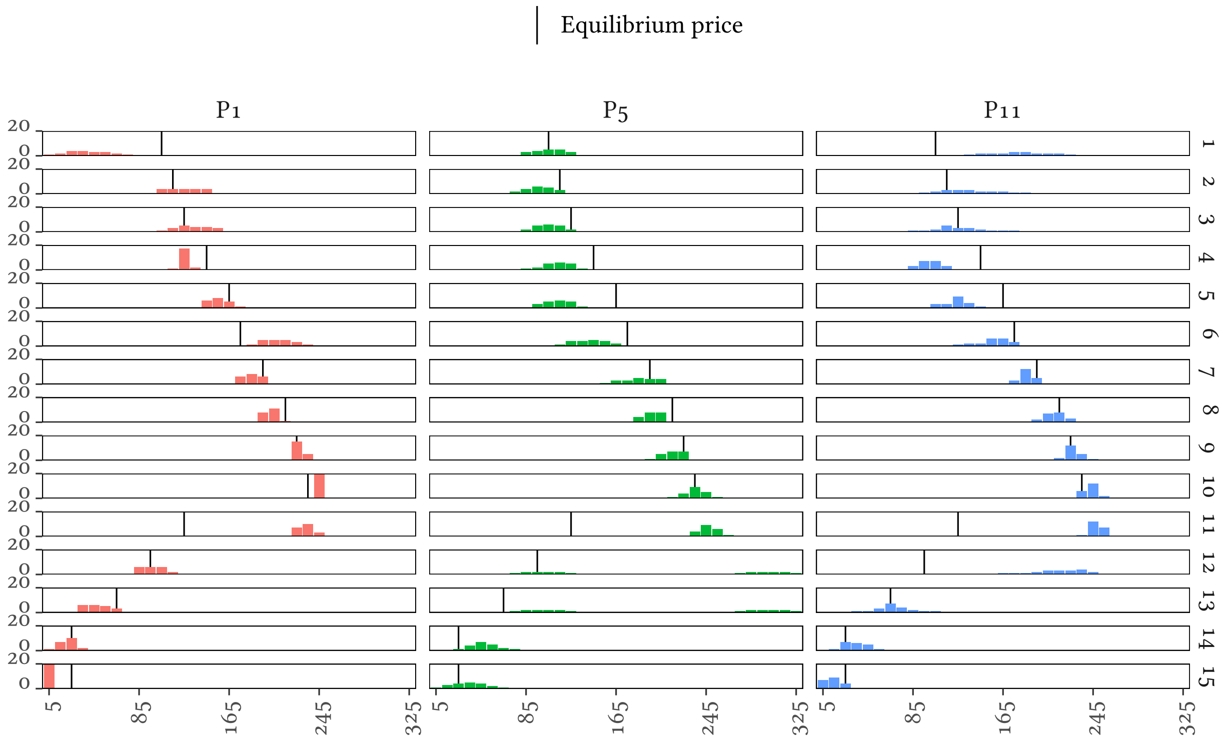


Figure 1.5.: An example of realized one-period-ahead price forecasts from the experiment. Each column displays a particular trader's forecasts for each of the 15 rounds in the market. Realized prices as well as the traders' maximum buy price and minimum sell price are also shown.

three different participants in a particular market. These participants are P1, P5, and P11[17]. In the top line the traders' first-round predictions are shown, and in the bottom line their final predictions are shown. Each of the small graphs shows price bins on the primary axis and the number of points (out of 20, each equivalent to 5%) allocated to the bins on the secondary axis. For instance, we see that P1 has allocated all 20 points to a single bin in round 10. The graph also shows the realized equilibrium price with the vertical, brown line. In this market a crash occurs in the 11th period.

Wider distributions, such as the ones seen in the first period, suggest that the trader is more uncertain about the price. Traders tend to make tighter distributions, concentrating on only a few potential prices, before period 11 – the crash. Note that each of the three traders behaves differently after the crash. P1 immediately switches to the new equilibrium price trajectory. P5 uses a bimodal distribution, putting some weight on the return to the bubble price and some weight on the fundamental value. P11 assumes the crash was a temporary shock, and expects the prices will return to previous levels—although with less certainty, as evidenced

17 While we only show three traders here, the data provides the full set of expectations for each participant for each round.

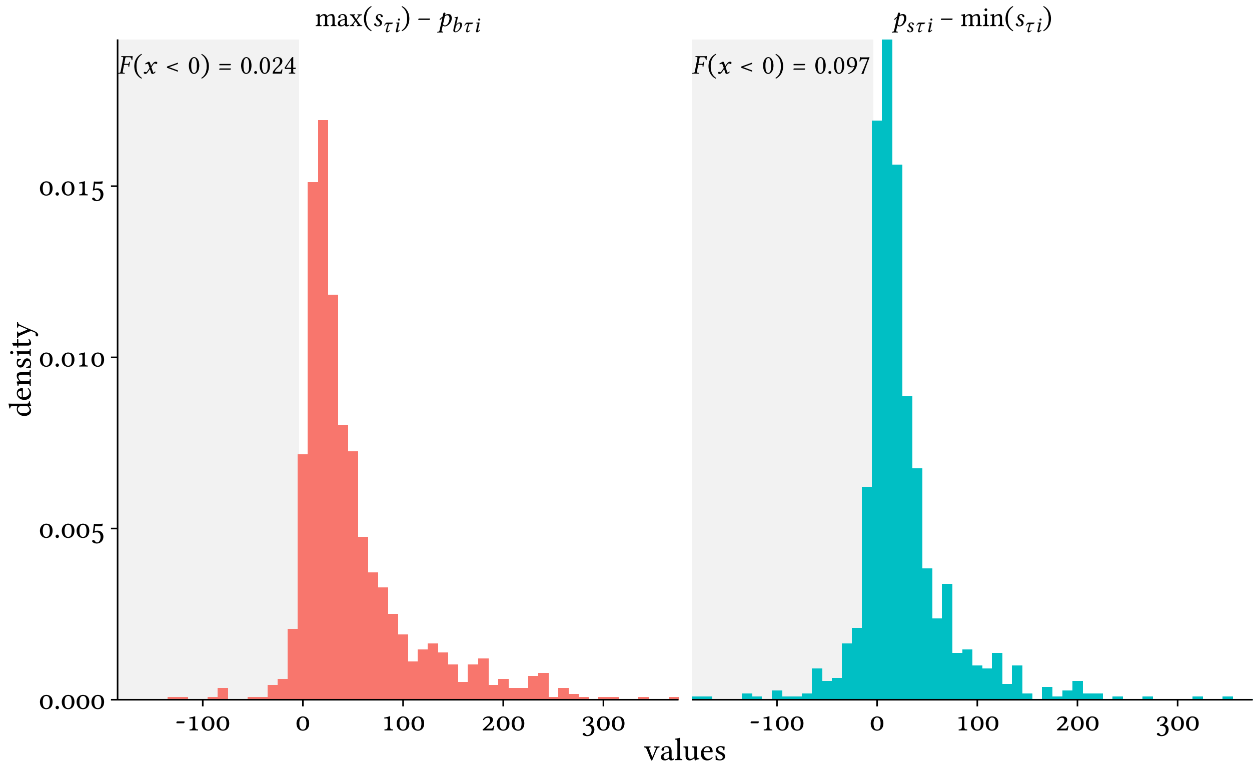


Figure 1.6.: The distributions of difference between bids $p_{b\tau}$ and asks $p_{s\tau}$ and forecasts $s_{i\tau}$ across all traders i and periods τ.

by the variance of the forecast.

A simple check to verify that the participants understood the game and acted rationally is to analyze whether forecasts are in line with their bids and asks. For instance, if a trader expects prices to be between 100 and 140, he should not offer to buy shares at a price above 140. Figure 1.6 shows the density of the difference between the maximum forecast and the bid on the left, and the difference between the ask and the minimum forecast on the right, over all rounds τ and all traders i. If traders are rational, they should not bid at a price above their expectations.[18]

If all traders acted rationally, there should be no support for negative values in the distribution of $\max(s_{\tau i}) - p_{b\tau i}$ shown in Figure 1.6. As the graph shows, most of the support is on positive numbers, suggesting that traders mostly understand the tasks. Only 2.4% of the bids are incompatible with the forecasts, in the sense that people are willing to buy at higher prices than what they expect. The graph also shows that 9.7% of the asks are lower than the minimum price the trader expects. This is, however, less of a concern, as the ask price only reflects the lowest price they are willing to accept.

18 We cannot make an equally strong statement for asks, as the optimal ask strategy depends on trader i's expectations about all other traders' beliefs.

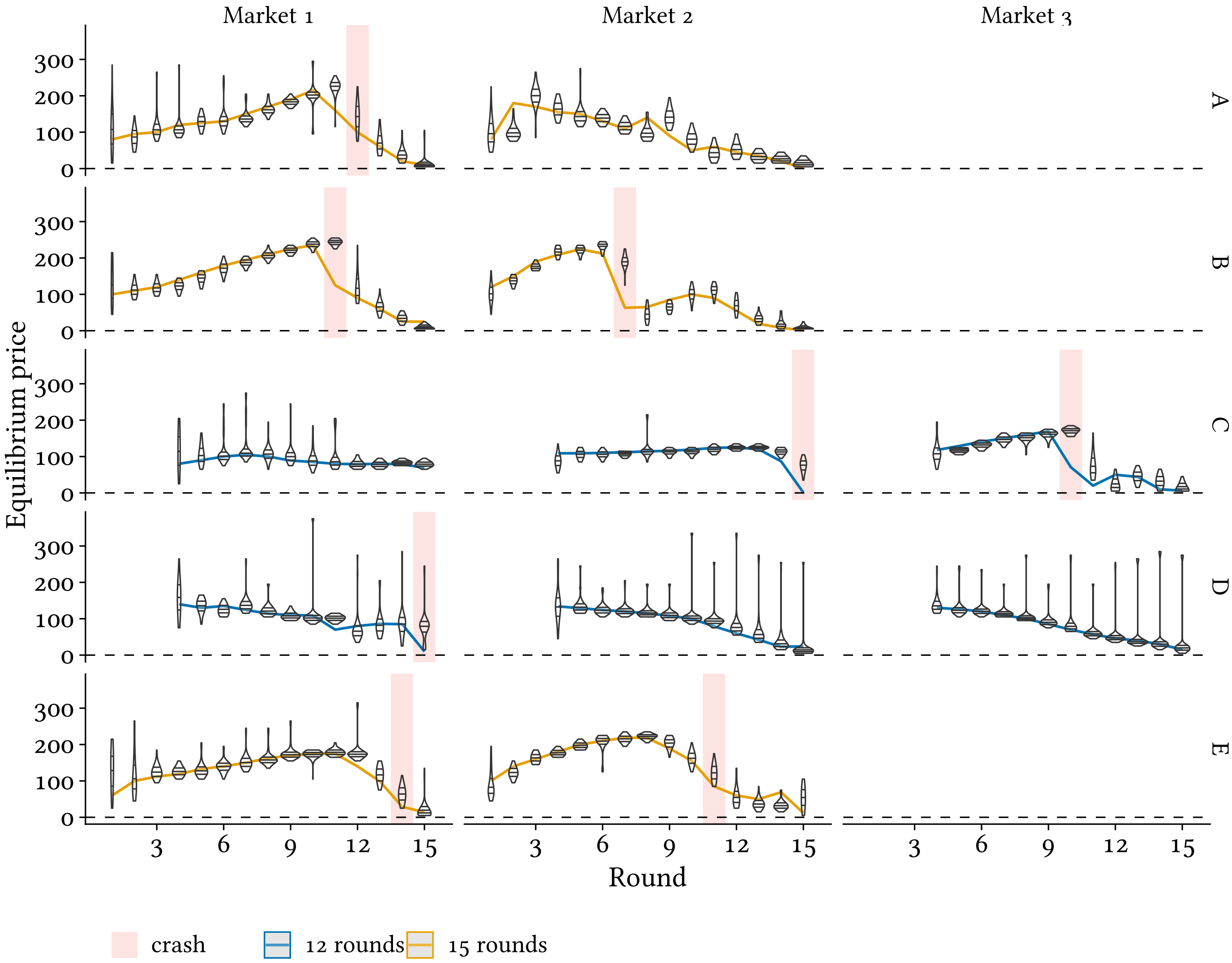


Figure 1.7.: Price expectations and realized prices. The lines show the development of the price. The violin distributions show the expectations of all participants, including the upper and the lower quartile, as well as the median, as shown by the horizontal lines.

Finally, Figure 1.7 shows the market certainty and equilibrium prices. Market certainty is displayed using a violin plot where the width of the violin shows the probability mass at a given price. Here, the market expectation is simply the union of all forecasts by all traders, treated as a single distribution.

Some patterns are clear across sessions. In general, market expectations are more dispersed in the first round, as shown by the wider range between the upper and the lower quartiles. This also seems to be the case across sessions. However, the expectations in the first round of the second market tend to be more certain than in the first market. This can be seen across all sessions. In the interim periods between the beginning and the crash, market expectations tend to become more certain, as measured by smaller distances between the upper quartiles and the lower quartiles. We see this, for example, in the first market of sessions A and B. Crashes are not widely expected, and the realized price in a crash tends to be lower than the lower quartile of expectations. This is the case for all the crashes observed in the data. Interestingly, the standard deviation of market expectations is sometimes increasing ahead of a crash—for instance in Session E. At other times, the crash is completely unexpected, as is the case of session B.

1.5. Results

1.5.1. Predictions and forecast

We turn attention to Hypothesis 1. This hypothesis states that we expect traders to narrow their beliefs after successfully predicting prices. In other words, success at predicting will lead to more certain beliefs, expressed as a tighter belief distribution. A longer strike of correct predictions for traders can itself lead to a longer bubble, under this hypothesis. Potentially, narrower beliefs, expressed through tighter forecast distributions, would lead to longer bubbles—as traders are essentially *riding the bubble*.

To investigate the claim, we look at whether increased success leads to a tighter forecast distribution. We expect that if a trader has a high degree of success in his forecast in period $t-1$, as measured by his forecast scoring, S_{t-1}, he will be more likely to predict a tighter distribution, i.e. a distribution with a smaller σ. We can express this using the following equation,

$$\sigma_{it} = \beta S_{it-1} + \alpha' \lambda + \epsilon_{ti}.$$

Here, λ represents a set of dummy variables, such as the trader ID, the round number, and market and session numbers. If traders become more certain as they make correct predictions the sign on β should be negative. The regression results are shown in Table 1.6. The data

Table 1.6.: Regressions on traders' certainty and success in forecasting using σ as the dependent variable. As the regressions show, traders become more certain as they successfully forecast the movement.

	σ_{it}				
	(I)	(II)	(III)	(IV)	(V)
S_{it-1}	-0.06^{***}	-0.05^{***}	-0.04^{***}	-0.04^{***}	-0.04^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Session FE		✓	✓	✓	✓
Participant FE			✓	✓	✓
Market FE				✓	✓
Round FE					✓
Num. obs.	1800	1800	1800	1800	1800
R^2	0.03	0.07	0.33	0.34	0.36
Adj. R^2	0.03	0.07	0.31	0.32	0.33

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Robust std. errors.

consists of all traders who participated in the experiment.

In the first column of Table 1.6 we show the results of the regression using the lagged forecast score S_{it-1} on the standard deviation of the forecast in period t. Note that the intercept is included in the regression, but is not reported. In the first column we show the simplest regression of S_{it-1} on $\sigma_i t$. There is a clear and stable negative relationship between the two variables, suggesting that increased prediction success in the previous round leads to a higher level of certainty, as measured by the width of the forecast distribution. Using the result of regression (I), a one-unit increase in lagged forecast earnings leads to a 0.06 reduction of the standard deviation.

The regressions displayed in Table 1.6 thus suggest that an increase in the forecast performance of one standard deviation would lead to a reduction of the standard deviation of σ_{ti} of between 0.19, in regression (I), and 0.15, in regression (V). The regressions vary in the included fixed effects, controlling for specific factors affecting each session (i.e, the time of day when it was held), market (learning effects) or participant (IQ or risk attitude). In Table 1.7 we present results where we use the range of participants' beliefs instead of dispersion, and the results hold.

This suggests a strong behavioral pattern that, in combination with our other hypothesis, can explain the amplification of bubbles.

Table 1.7.: Regressions on traders' certainty and success in forecasting using the range of the forecast as a dependent variable.

	range$_{it}$				
	(I)	(II)	(III)	(IV)	(V)
S_{it-1}	-0.18^{***}	-0.15^{***}	-0.14^{***}	-0.13^{***}	-0.13^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Session FE		✓	✓	✓	✓
Participant FE			✓	✓	✓
Market FE				✓	✓
Round FE					✓
Num. obs.	1800	1800	1800	1800	1800
R^2	0.04	0.08	0.34	0.35	0.38
Adj. R^2	0.04	0.07	0.32	0.33	0.36

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Robust std. errors.

1.5.2. Convergence of expectations

We now turn to Hypothesis 2. The hypothesis states that while prices and expectations converge in means, traders will still display heterogeneous beliefs. To investigate this issue we use the concordance correlation coefficient (ρ_c) proposed by Lin (1989). This coefficient, ρ_c for two distributions X_1 and X_2, can be calculated as

$$\rho_c(X_1, X_2) = \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 + \left(\mu_1 - \mu_2\right)^2}$$

where μ_i is the mean of X_i, σ_{ij} is the covariance between X_i and X_j, and σ_i^2 is the variance of X_i (Lin, 1989).[19]

This coefficient is in $[-1, 1]$, and measures the extent to which pairs of observations from the two distributions fall on a 45°-line. A value of -1 can be interpreted as perfect disagreement whereas a value of 1 can be interpreted as perfect agreement.

As simple starting point to test Hypothesis 2, we analyze whether the observed concordance correlation coefficient is increasing over time. For each round and for each pair of traders i, j we calculated ρ_c. The results are shown in Table 1.8, where the median concordance correlation coefficient is calculated for pairs of traders for every period. In parentheses we

19 The concordance correlation coefficient relates to the Pearson correlation coefficient ρ and Lin shows that ρ_c poses the following characteristics: $-1 \leq -\left|\rho\right| \leq \left|\rho_c\right| \leq \left|\rho\right| \leq 1$; $0 = \rho \iff \rho_c = 0$; $\rho_c = p \iff (\mu_1, \sigma_1) = (\mu_2, \sigma_2)$; and that $\left|\rho_c\right| = 1$ only occurs for two samples in perfect agreement or disagreement. See Lin (1989) for a more detailed treatment of the characteristics.

Table 1.8.: Average concordance correlation coefficient, ρ_c between rounds. Plain numbers denote the mean, while numbers in parentheses denote the standard deviation. The superscripted stars denote the significance of the p-value of the Wilcoxon Signed Rank Sum Test between the observations from period j in market i and $i-1$.

t	1		2		3	
1	0.246	(0.236)	0.334***	(0.275)		
2	0.449	(0.281)	0.451	(0.251)		
3	0.475	(0.264)	0.523**	(0.274)		
4	0.384	(0.289)	0.446***	(0.282)	0.437*	(0.266)
5	0.440	(0.261)	0.517***	(0.262)	0.566**	(0.240)
6	0.457	(0.280)	0.496**	(0.278)	0.510	(0.290)
7	0.447	(0.294)	0.524***	(0.265)	0.428	(0.297)
8	0.448	(0.251)	0.419	(0.294)	0.457	(0.306)
9	0.459	(0.255)	0.433	(0.286)	0.469	(0.299)
10	0.422	(0.294)	0.442	(0.272)	0.383	(0.266)
11	0.433	(0.287)	0.444	(0.276)	0.364	(0.314)
12	0.360	(0.280)	0.417***	(0.280)	0.444	(0.304)
13	0.404	(0.275)	0.484***	(0.289)	0.530***	(0.342)
14	0.393	(0.278)	0.438**	(0.265)	0.461	(0.297)
15	0.282	(0.291)	0.256	(0.276)	0.318	(0.294)
All	0.407	(0.282)	0.442***	(0.284)	0.447	(0.301)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

show the standard deviation of the concordance correlation coefficient. The first row shows that the median concordance score in market 1, round 1 is 0.175 and that it goes up to 0.214 by the second market. In addition, the standard deviation slightly increases in the second market. The stars summarize the significance of the Wilcoxon Signed Rank Sum Test with a one-sided alternative hypothesis.[20] Specifically, for a pair of players i and j, we test the null that $\rho_{c,ij}^{m,r} - \rho_{c,ij}^{m+1,r}$ is zero where m and r denote the market and period, respectively. The difference between column 1 and column 2 is therefore of interest to address Hypothesis 2. There is a moderate increase in ρ_c in round 1 between market 1 and market 2[21]. In general, the level of agreement tends to go up between market 1 and market 2, as shown by the bottom line of the table. However, the effects vary from round to round, e.g. round eight we cannot reject the null that there is no difference between the level of agreement.

In Appendix A.1.2 we carry out the same analysis using the Bhattacharyya coefficient. The Bhattacharyya coefficient serves as a good alternative, as it enables comparison of distributions without taking the specific moments into account. The specific measure does not, however,

20 We use a non-parametric test over a t-test, as the distributions of differences between values of market 1 and market 2 have fat tails and are only spread over the domain $[-1, 1]$.

21 There is a larger increase still between market 2 and 3 (the first round of market 3 is denoted 3 in the table).

Table 1.9.: Regression analysis on overlap of expectations.

	$\rho_{c,ij} - \bar{\rho}_{c,ij}$				
	(I)	(II)	(III)	(IV)	(V)
Market 2	0.035^{***}	0.027^{**}	0.027^{**}	0.028^{**}	0.028^{**}
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Market 3	0.016	−0.008	−0.003	0.000	0.000
	(0.010)	(0.012)	(0.013)	(0.013)	(0.013)
$p_{t-1} - f_t$		-0.000^{***}	-0.000^{***}	−0.000	−0.000
		(0.000)	(0.000)	(0.000)	(0.000)
Session FE			✓	✓	✓
Round FE				✓	✓
Participant FE					✓
Num. obs.	10692	9900	9900	9900	9900
R^2	0.004	0.007	0.008	0.060	0.061
Adj. R^2	0.004	0.007	0.007	0.058	0.054

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Std. errors clustered on IDs.

alter the conclusions.

To test Hypothesis 2 in a more structured way, we test whether expectations become more aligned. To do this we regress the concordance correlation coefficient on market fixed effects to see if the there is a convergence effect over time. We include the fixed effects to make sure that other unobserved factors are not driving the results. The regressions are shown in Table 1.9, where we regress the within group (defined by pairs of traders) difference between agreement and average agreement. Standard errors are clustered on groups of traders to allow for within-group heteroscedasticity (e.g. Abadie et al., 2017). This may arise due to learning or other time-varying factors.

Our preferred specification is the first regression, (I), which simply shows that the rate of agreement is increasing by, on average, 0.035 between the first and the second market. For this to be a consistent estimator we have to assume that any unobservables that would affect agreement are fixed over time and thus are removed by the session fixed effect. Nonetheless, we include a number of other fixed effects and show that the results are similar under other specifications with more fixed effects. In column (II)–(V) we include the lagged bias, defined as the difference between the realized price and the fundamental value, to capture the variation in market prices across markets and sessions. Columns (III)–(V) show that the result is robust to the inclusion of a number of fixed effects that capture time-invariant effects. We use the fixed effects to remove any within effects within sessions (II), rounds (III) or groups of traders (IV, V).

Table 1.10.: Regression analysis of difference in expectations of types.

	$\mathrm{E}_{i,t}\{p_t\} - E_{-i,t}\{p_t\}$				
	(I)	(II)	(III)	(IV)	(V)
neither seller nor buyer	−1.143	−2.073	−2.115	−2.546	−2.565
	(1.970)	(1.681)	(1.691)	(1.696)	(1.686)
seller	−4.925**	−7.144***	−7.218***	−7.429***	−7.027***
	(1.891)	(1.626)	(1.632)	(1.634)	(1.619)
$p_{t-1} - f_{t-1}$		0.038**	0.043**	0.045**	0.099***
		(0.013)	(0.014)	(0.014)	(0.019)
$\mathrm{E}_{i,t-1}\{p_{t-1}\} - p_{t-1}$		0.450***	0.452***	0.457***	0.492***
		(0.017)	(0.017)	(0.017)	(0.018)
Session FE			✓	✓	✓
Market FE				✓	✓
Round FE					✓
Num. obs.	1944	1800	1800	1800	1800
R^2	0.004	0.276	0.277	0.280	0.301
Adj. R^2	0.003	0.275	0.274	0.276	0.292

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Robust std. errors.

As Table 1.9 show the regression results support the initial finding that there is an increase in the level of agreement (in the sense of the concordance correlation coefficient) between session one and session two.[22]

Finally, we investigate whether the heterogeneity in beliefs that is still observable can be explained by trader types. As Figure 1.7 shows, beliefs are exhibiting some heterogeneity, even in the later markets. In Figure 1.8 we split the expectations into buyers and sellers. *Buyers* are either realized buyers or traders who only make buy orders in a given round while *sellers* are defined to be either realized sellers or traders who only make sell orders in a given round. Each pair of bars denotes the range of expectations among a given group. Figure A.1 in the appendix shows the same plot for individual buyers and sellers.

While the graph suggests that beliefs may differ in some periods, like the burst of the bubble in session A, it is hard to tell if there is a persistent effect. To investigate this further, we perform the following baseline regression

$$p^e_{t,i} - p^e_t = \beta_0 + \beta_1 \mathrm{seller}_{i,t} + \beta_2 \mathrm{neither}_{i,t} + \epsilon_{i,t}$$

Results are in Table 1.10. The left-hand-side variable is the difference between trader i's belief

22 The regression results also suggest that we cannot rule out that there is no increase in the agreement between market 2 and market 3. These results are also robust dropping the observations of sessions, with 12 market rounds

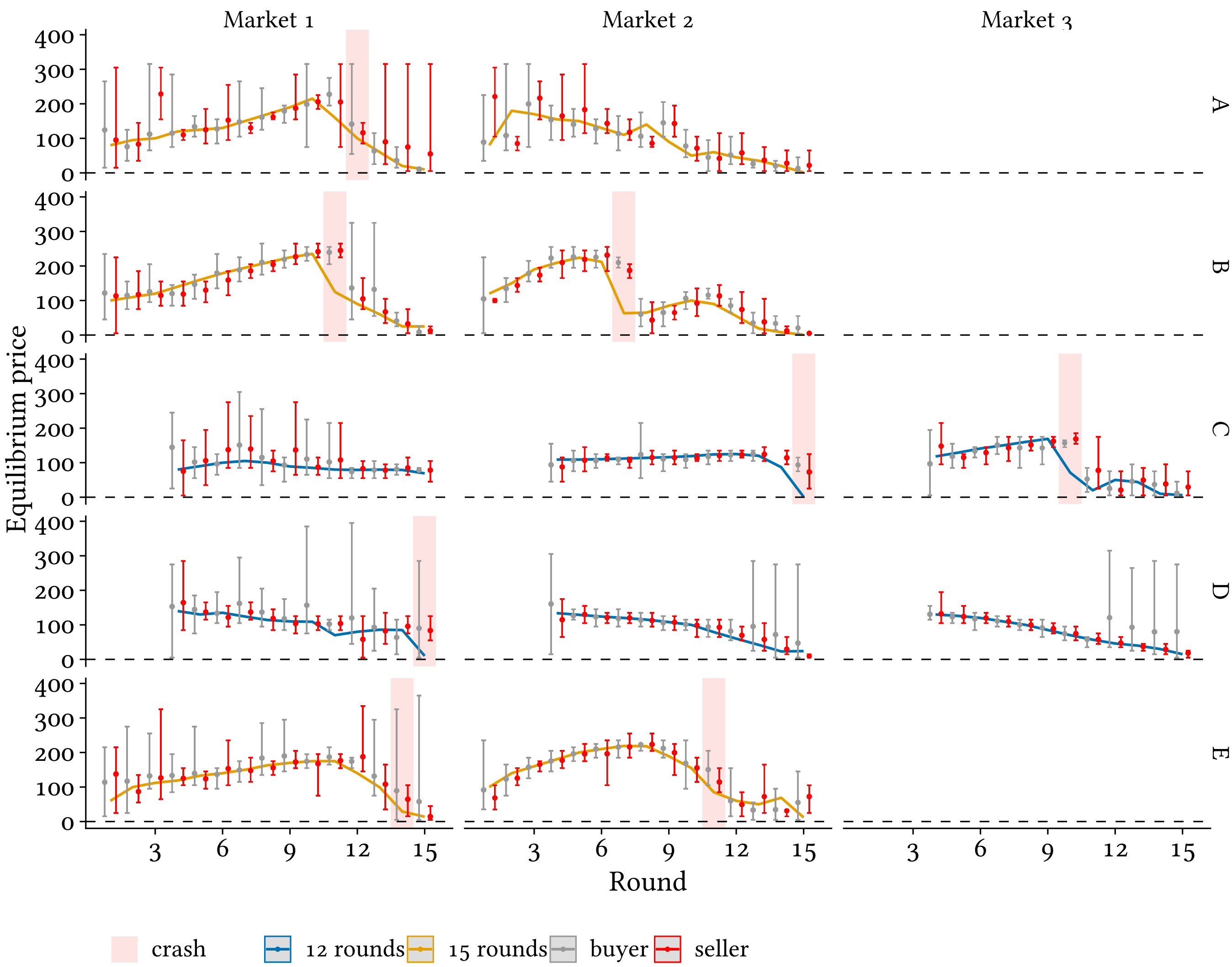


Figure 1.8.: Price expectations of buyers and sellers. Each trader is categorized as a *buyer*, a *seller* or *neither*. The dots denote the mean expectation of each group while the bars show the range of beliefs.

about the price development in a specific round and the mean belief in that round. The key hight hand side variable is the type of the trade. The baseline is buyers, and the two other types, *neither a buyer nor a seller* and *seller*, are shown. In column (II)–(V) we include the lagged bias, $p_{t-1} - f_{t-1}$; i.e. the difference between the realized price and the fundamental value. We also include the lagged forecast bias, $E_{i,t-1}\{p_{t-1}\} - p_{t-1}$, i.e, the difference between the expected price and the realized price.

We include these variables to capture the state of a particular round, as they may have an impact on the beliefs of traders. We also include a number of fixed effects to capture variations across sessions, markets and market rounds.[23]

23 For the regressions to be consistent we must assume that there are no systematically variations that are correlated with the regressors and the left-hand variable that is not included.

Table 1.11.: Regression analysis of difference in expectations of types.

	$E_{i,t}\{p_t\} - E_{-i,t}\{p_t\}$				
	(I)	(II)	(III)	(IV)	(V)
neither seller nor buyer	0.372	−2.187	−2.513	−2.984	−3.118
	(2.187)	(1.798)	(1.802)	(1.806)	(1.803)
seller	-5.856^{**}	-6.921^{***}	-6.781^{***}	-7.094^{***}	-6.670^{***}
	(2.093)	(1.735)	(1.739)	(1.739)	(1.735)
$p_{t-1} - f_{t-1}$		0.043^{***}	0.040^{**}	0.040^{**}	0.095^{***}
		(0.013)	(0.014)	(0.014)	(0.022)
$E_{i,t-1}\{p_{t-1}\} - p_{t-1}$		0.538^{***}	0.546^{***}	0.551^{***}	0.574^{***}
		(0.020)	(0.020)	(0.020)	(0.020)
Session FE			✓	✓	✓
Market FE				✓	✓
Round FE					✓
Num. obs.	1644	1500	1500	1500	1500
R^2	0.006	0.335	0.339	0.343	0.356
Adj. R^2	0.005	0.333	0.336	0.339	0.346

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Robust std. errors.

Our preferred specification in Table 1.10 is (II). This regression shows that, on average, sellers tend to have lower mean expectations about the realized price than buyers, by 6.5 points. We include controls for the lagged bias, which is an observed indicator for the state of the market that all traders observed. Higher lagged values of both the price bias and the forecast bias both have positive coefficients. We include the variables to control for the state of the particular market round to gain a consistent estimate.

The results are not driven by the inclusion of post-crash rounds. Table 1.11 shows the same regressions as Table 1.10, but using only data in up to the crash. As Table 1.11 shows, the results are not significantly altered by the exclusion of post-crash data.

In combination, the results presented in this section show that while agreement is increasing over time, there is still significant heterogeneity among and across traders. On average, buyers tend to have somewhat higher expectations about the price development compared with sellers as shown in Table 1.10. An unusual feature that can be observed in some sessions is that when buyers expectations fail to adapt quickly to the fundamental value, there is not a crash in the traditional sense, but a slow fall of bubbles.

1.5.3. Market volatility and initial expectations

We now turn to the related question of whether higher initial disagreement in forecasts leads to smaller bubbles, as proposed by Hypothesis 3. The hypothesis states that in a market where traders agree more, there is an increase likelihood of them "riding the bubble." If this is the case, an initial level of agreement in the market would lead to a larger bubble. A potential explanation would be that people form wrong expectations about the equilibrium price-path, disregarding the fundamental value. Alternatively, people could be betting on riding the bubble and leaving it in time before it bursts.

To investigate Hypothesis 3 we use market-level variables. This allows us to take a broad approach to problem. However, it also implies fewer degrees of freedom. We use the two bubble measures introduced in Stöckl et al. (2010) and used in Kirchler et al. (2012). These measures are the measure for relative absolute deviation (*RAD*) and relative deviation (*RD*), also shown in Table 1.5. We use these measures as they are insensitive to the choice of parameters, such as the number of rounds (Stöckl et al., 2010). Due to the differences between markets lengths across sessions, this is a desirable property for the test.

RAD and *RD* measures are defined as,

$$RAD = \frac{1}{N}\sum_{t=1}^{T}\frac{|p_t - f_t|}{|\bar{f_t}|}$$

$$RD = \frac{1}{N}\sum_{t=1}^{T}\frac{p_t - f_t}{\bar{f_t}}$$

Both measures compare the realized price p_t to the fundamental value f_t. As it is summed over all periods, the duration of the bubble and its size lead to a larger value of the measurement. The sum is divided by the average fundamental value, $\bar{f_t}$. This scales the numerator to make the sessions with different durations of markets comparable. *RAD* measures the average level of mispricing, be it a bubble or a burst. This means that it is able to capture both the undervaluation (overvaluation) when p_t is smaller (bigger) than f_t.[24] *RAD* is similar to the amplitude measure, as it measures overall deviation. *RD* does not use the absolute value, and, as such, positive and negative values can offset each other.[25]

As our explanatory variable we use the concordance index ρ_c, introduced above in Equation (A.1). As the analysis is conducted on a market level, we use the average value over the first τ periods $\bar{\rho}_{c,\tau} = \frac{1}{\tau}\sum_{t=1}^{\tau}\frac{1}{N(N-1)}\sum_{i,j}\rho_{c,ij,t}$, where i and j are traders and N is the number of participants.[26] As shown in the previous section, $\bar{\rho_{c,\tau}}$ increases with experience. We use fixed effects for the market and the sessions to control for this.

24 This captures the undervaluation in initial periods.

25 In our sample, all markets except market 2 in session A and market 3 in session D featured positive relative deviations.

26 We also vary τ as a robustness check.

Table 1.12.: Results of Regression (1.2).

	RAD			*RD*		
	(I)	(II)	(III)	(IV)	(V)	(VI)
(Intercept)	0.38			0.05		
	(0.32)			(0.23)		
ρ_c	0.01	1.42	1.39*	0.34	1.28*	1.70*
	(0.74)	(0.70)	(0.37)	(0.54)	(0.55)	(0.43)
Market FE		✓	✓		✓	✓
Session FE			✓			✓
Num. obs.	12	12	12	12	12	12
R^2 (full model)	0.00	0.56	0.96	0.04	0.52	0.89
Adj. R^2 (full model)	-0.10	0.40	0.88	-0.06	0.34	0.71

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Robust std. errors.

We specify the following regression,

$$y_{s,m} = \beta\bar{\rho}_{c,\tau} + \lambda_m + \lambda_s \tag{1.2}$$

where the dependent variable $y_{s,m}$ is the bubble size variable (i.e, *RAD* or *RD*), and λ_m and λ_s are market and session fixed effects respectively. The results are shown in Table 1.12 where we used $\bar{\rho}_{c,2}$ for sessions C and D (with 12 rounds) and $\bar{\rho}_{c,3}$ for sessions A, B and E (with 15 rounds). We use different values of τ to keep a similar fraction of rounds for the measure ρ_c.

Even though the results are not statistically strong due to the low number of observations, we find support for the hypothesis that more disagreement dampens the size of the bubble. A large $\bar{\rho}_{c,\tau}$ means a large overlap of traders' expectations, and this implies a larger bubble coefficient. This provides support for the Minsky hypothesis, in the sense that more stable expectations tend to cause bigger market crashes.

1.5.4. Explanatory power of certainty on price formation

We now turn to the Hypothesis 4. We are interested in the predictive power of trader certainty on future price changes. In other words, we want to examine to what extent the higher moments of traders' beliefs can be used to understand and predict price movement in the experimental market.

We start by using the most complete specification used in Haruvy et al. (2007) to explain the relationship between the change in price and various variables that may affect it. The specification used is

$$\Delta p_{m,t} = \beta_0 + \beta_1(p_{m,t-1} - f_{t-1}) + \beta_2 \Delta p_{m,t-1} + \beta_3 N_{m,t} + \beta_4 \Delta E_{m,t} + \beta_5(p_{m,t-1} - E_{t-1}) + \beta_6 \sigma^m_{m,t} + \epsilon_{m,t}. \tag{1.3}$$

The first term, $p_{m,t-1} - f_{t-1}$ is the lagged bias, and shows the difference between the realized price and the fundamental value of the asset in the previous period. The second term, $\Delta p_{m,t-1}$ represents the lagged price difference. The third term, $N_{m,t}$ corresponds to the number of short-term pessimistic traders, as defined by Haruvy et al. (2007). This number is calculated as the relative number of traders for whom the median forecast in period t is lower than the median of the forecast in period $t-1$.

We add three terms that measure further aspects of market certainty using our distributional information on traders' beliefs.

First, we use the change of the mean expectation from last period to this, $\Delta E_{m,t}$. This measures the change in the first moment of expectations. Second, we use the lagged difference between the realized price and the expectation, $(p_{m,t-1} - E_{t-1})$. This works as an error correction term that measures the mistakes in traders' expectations in the last period. Finally, we add a measure of dispersion of beliefs or certainty.

We use two measures of dispersion of beliefs. At the individual level we use (σ_t), and at the market level (σ^m_t). For the individual level, we calculate the standard deviation of beliefs to obtain a single measure of certainty for each period. For the aggregate level we create a *market certainty* variable, $\sigma_{m,t}$. We combine the forecasts of all traders before calculating the statistics. This allows us, unlike the literature, to obtain the intensity of market beliefs. In particular, we create a new forecast probability mass function containing all forecasts from all traders and calculate the statistics based on this forecast. We interpret this as "market sentiment".

To illustrate, consider a market with two traders, where each trader creates a forecast distribution consisting of three points. If the first trader makes the prediction $\{1/3, 1/3, 1/3\}$, that is the trader believe that the price will be either p_1, p_2 or p_3 with $\frac{1}{3}$ chance, and the second trader makes the prediction $\{1/2, 1/4, 1/4\}$, we calculate the statistics on the probability mass function $\{5/12, 14/24, 14/24\}$. This approach only makes a difference for non-linear statistics, such as the variance. For most of the variables, this has little impact. Likewise, one can use the median rather than the mean in (1.3).[27]

Results using Regression (1.3) and the augmentation factors are shown in Table 1.13. All regressions use the change in price Δp_t as dependent variable. From left to right, each column displays a more complex model, following Haruvy et al. (2007, equation (7)). From an econometric point of view, the efficiency of the regressor depends on what assumptions are made about the underlying process and the inter-dependencies of the variables.

27 This does not alter the results.

Table 1.13.: Regression results of (1.3). Note that the coefficient on the intercept is omitted.

	Δp_t						
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Δp_{t-1}	0.34^{***}	0.00	0.31^{***}	-0.43^{*}	-0.42^{*}	-0.37	-0.42
	(0.07)	(0.10)	(0.07)	(0.21)	(0.20)	(0.19)	(0.22)
$p_{t-1} - f_{t-1}$	-0.28^{***}	-0.31^{***}	-0.29^{***}	-0.26^{***}	-0.28^{***}	-0.22^{***}	-0.27^{***}
	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.06)
N_t		-34.19^{***}			-25.37^{**}	-11.85	-8.52
		(7.08)			(7.65)	(7.87)	(8.88)
σ_t^m			-0.26^{*}	-0.30^{**}	-0.20	-0.42^{***}	-0.62^{***}
			(0.11)	(0.11)	(0.11)	(0.11)	(0.14)
ΔE_t				0.75^{***}	0.50^{*}	1.25^{***}	1.49^{***}
				(0.20)	(0.21)	(0.26)	(0.28)
$E_{t-1} - p_{t-1}$						-0.93^{***}	-1.09^{***}
						(0.22)	(0.24)
Session FE							✓
Market FE							✓
Round FE							✓
Num. obs.	138	138	138	138	138	138	138
R^2	0.29	0.39	0.32	0.38	0.43	0.50	0.61
Adj. R^2	0.28	0.38	0.30	0.36	0.40	0.47	0.53

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ Robust std. errors.

The first regression, (I) shows that both the magnitude of the change in the previous round, Δp_{t-1}, and the realized necessary correction to the fundamental value, $p_{t-1} - f_{t-1}$, are significant. While the coefficient on Δp_{t-1} becomes insignificant and even changes sign as more variables are added, the coefficient on $p_{t-1} - f_{t-1}$ remains negative and of approximately the same magnitude throughout all regressions. This suggests that, as the difference between the equilibrium price and the fundamental value grows larger, the change in price is dampened. In other words, the fundamental price does play a role in how the market moves as a whole. For this estimation to be consistent we have to assume that there are no variables that are left out that are correlated with both the dependent variable and the included right-hand-side variables.

Column (II) replicates the Haruvy et al., 2007 regressions with economically similar results. It shows that the number of pessimistic traders is an effective measure to predict the change in prices, and as such it can be seen as a measure of market certainty. [28]

28 Unlike Haruvy et al. (2007) the measure used here is divided by the number of participants to normalize it, which explains why the coefficient estimate is larger than the one obtained by Haruvy et al. (2007).

In Haruvy et al. (2007) the distribution of beliefs by participants is symmetric and, as such, the mean and the median coincide. The results shown in Table 1.13 do not change when we use the mean of the forecasts rather than the median to define N_t. Nor does it change when we require both the mean and the median of round t to be lower than those of round $t-1$. For regression (II) to be unbiased it is required that there are no omitted systematic factors that are correlated with both the regressors and the dependent variable.

In regression (III) we add σ_t^m, the standard variance of the market forecast, which we denote *market uncertainty*. The coefficient is negative and significantly different from zero across specifications (III)–(VII). The fact that the coefficient is negative suggests that a larger spread of the forecast distribution leads to less price movement. In other words, as market uncertainty about the price increases, the price movement is dampened. As more controls are added in regression (VI) and (VII), the effect of market uncertainty on price movement increases.

In Regression (IV), (V) and (VI) we include measures of the two first moments of expectations. Our preferred specification is (VI).[29] Regression (V) also includes the N_t measure of pessimistic traders defined by HARUVY07. Note that the inclusion of N_t in regression (V) leads to an insignificant coefficient on σ_t. However, once we include additional market sentiment measures, namely the forecast bias $E_{t-1} - p_{t-1}$, and the change in the mean expectations ΔE_{t-1}, the coefficient on σ_t^m is significant again, as shown in regression (VI).

These regressions essentially test Hypothesis 4, as we include information on the market certainty. Our hypothesis states that information is indicative of price movements. The simplest form of testing this is through regression (IV), where we include the change in the mean expectation as well as the dispersion of beliefs σ_t. As the regression shows, the inclusion of these two variables is highly significant in explaining the change in the price level, and more so than Δp_{t-1}.

Regression (VI) includes an additional variable, forecast bias $E_{t-1} - p_{t-1}$. The coefficient on this variable captures the reaction when expectations are not aligned with realized prices. Likewise, the change in the first moment of the expectation, ΔE_t, follows the same sign as the change in price, indicating that traders are usually capable of predicting the direction of the price movement. Importantly, we find that once we control for the market sentiment variables, the number of pessimistic traders is no longer statistically significant.

For regressions (IV), (V) and (VI) to be consistent and efficient it has to be (a) that market certainty matters for realized prices and (b) the market certainty measure adequately captures individual-level heterogeneity.

In the final regression, (VII) we add a number of fixed effects to control for the particular

29 We use a robust estimation of the variance-covariance matrix to allow for heteroskedasticity as there may be differences in the variance-covariance structure between sessions and markets.

Table 1.14.: Regression results of (1.3), using only observations from before a crash.

	Δp_t						
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Δp_{t-1}	0.37^{***}	-0.06	0.32^{***}	-0.21	-0.27	-0.31	-0.36
	(0.08)	(0.11)	(0.08)	(0.19)	(0.18)	(0.16)	(0.21)
$p_{t-1} - f_{t-1}$	-0.10^{**}	-0.13^{***}	-0.13^{***}	-0.12^{***}	-0.14^{***}	-0.09^{**}	-0.10
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)
N_t		-28.46^{***}			-21.31^{***}	-11.09	-11.62
		(5.60)			(6.20)	(5.87)	(6.90)
σ_t^m			-0.25^{**}	-0.31^{***}	-0.20^{*}	-0.45^{***}	-0.44^{***}
			(0.08)	(0.08)	(0.08)	(0.09)	(0.11)
ΔE_t				0.52^{**}	0.29	1.11^{***}	1.06^{***}
				(0.17)	(0.18)	(0.23)	(0.25)
$E_{t-1} - p_{t-1}$						-0.99^{***}	-0.88^{***}
						(0.19)	(0.23)
Session FE							✓
Market FE							✓
Round FE							✓
Num. obs.	105	105	105	105	105	105	105
R^2	0.22	0.38	0.29	0.35	0.42	0.54	0.61
Adj. R^2	0.21	0.36	0.27	0.32	0.39	0.51	0.50

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ Robust std. errors.

session, market or the number of rounds. Adding the fixed effect does not alter the results significantly.

Note that if we let $\beta_4 = \beta_5 = \beta_6 = 0$ as in regression (II) we have a similar regression to the one used by Haruvy et al. (2007). Under Hypothesis 4 we should get $\beta_4 \neq 0 \wedge \beta_5 \neq 0 \wedge \beta_6 \neq 0$. In other words, there is additional explanatory power beyond the other variables displayed in (1.3). As can be seen from regressions (VI) and (VII), we can clearly reject that $\beta_4 = \beta_5 = \beta_6 = 0$. The only case where the coefficient on σ_t is not significant is in regression (v), but the sign and the magnitude remains in line with the other regressions, where the coefficient is significant.

As a robustness check we perform the same analysis using only data from before the crash and run the same regressions. This is done to rule out that the results are only driven by data observed after crashes. The results are shown in Table 1.14. The results are similar to those found for the full data-set in Table 1.13.

Overall the results in this section support Hypothesis 4. Table 1.13 shows that trader certainty has a strong predictive power on future price movements. When market certainty is high, prices tend to increase above the fundamental value.

1.6. Conclusion

Using an experimental asset market it is possible to shed light on the role of expectations in driving prices. We investigate the role of the dispersion of beliefs among traders in creating and amplifying asset bubbles. We embed a way of eliciting beliefs in the workhorse of experimental asset markets, Smith et al. (1988). The design allows us to avoid the problem of hedging and to capture arbitrary distributions of beliefs.

We find that traders' beliefs become less variable when their predictions are accurate, even if the price is above the fundamental value. Moreover, traders' expectations tend to converge as they become more experienced.

Importantly, we find that more agreement among traders causes larger bubbles. This sheds light on the Financial Instability Hypothesis. When traders' expectations are more scattered, the information is aggregated in the market in a way that dampens optimism. When expectations are more concentrated, market optimism seems to create larger bubbles. Finally, market certainty has predictive power over price fluctuations, beyond the mean expectation of the market. This has potential policy implications for the monitoring of financial markets that should be further explored.

Chapter 2.

The impact of macroeconomic uncertainty on demand: A study of demand for the US airline industry

2.1. Introduction

Abstract

A number of countries experienced increased economic uncertainty in the wake of the financial crisis. An important question is to what extent this uncertainty affect demand. To investigate this issue I examine empirically to what extent varying uncertainty affects the consumer demand for flight traffic using US micro demand data. I find that the elasticity of uncertainty is overall economically and statistically significant.

This paper examines to what extent demand for flight tickets is affected by macroeconomic events using a discrete choice model fed with recent microdata from the US flight market. In particular, I examine the impact of changes in the level of economic and political uncertainty on the demand for airline travel. This is an important research area, as it provides an insight into the dynamic effect of uncertainty on demand. This yields important insights into the effect of increased uncertainty caused by external events, from the consumers' point of view, such as the financial crisis.

Air transportation is interesting not only in its own right, but also for its properties. In particular: (i) airlines enable rapid transportation of goods and people, making broader scales of economics practical and allowing for vacations far from people's homes;; (ii) as air-transport tends to be relatively expensive and planned ahead one would be more inclined to believe that expected utility is able to adequately account for the observed behavior, and (iii) the primary service provided by airlines, namely transportation from point A to point B, is relatively simple, yet the level of service, connections, and local products vary from provider to provider and from region to region. Likewise, the valuation of the services and properties vary between people. The model is examined through demand estimation as performed in the industrial organization literature and discrete choice modeling in the spirit of, for instance, McFadden (1981) and Berry et al. (1995). The goals are: (i) to try to adequately account for the endogeneity that the data and the modeling limitation necessarily will introduce the decision-making framework, and (ii) to check the robustness of the product and market structure of air transport imposed by the econometrician.

The goal of the paper is to directly assert how increased uncertainty affect demand. To the extent that uncertainty is exogenous to the consumer and measurable, such an analysis is valid.

The paper proceeds as follows. Section 2.2 briefly examines the literature, focusing on previous empirical and theoretical studies. Section 2.3 displays a sketch structural model, using and elaborating on the models discussed in section 2.2. The destination and departures dataset used for the empirical estimation is discussed in section 2.4. In section 2.5 the results of the discrete choice model is presented and discussed. Simplifications are proposed and used to estimate the theoretical model. Finally, section 2.6 concludes and points toward future research.

2.2. Literature review

The consumer choice of airline transport has received much attention in the economics and engineering literature, typically modeled as a discrete choice. The discrete choice framework allows the researcher to estimate the probability of choice given covariates and the assumption that agents act so as to maximizing some objective function, typically a utility specification. However, while the estimation of the demand functions themselves has been extensively studied, the impact of the surrounding macroeconomic conditions is rarely considered.

A large body of literature has looked at the impact of uncertainty on the decision-making of economic agents (see Dixit and Pindyck, 1994, for a comprehensive overview). Bloom et al. (2007) and Bloom (2009) show that uncertainty matters both at a firm level and at an aggregated level. Likewise, Kellogg (2014) shows that in the Texas on-shore oil industry investment decisions are highly influenced by the uncertainty about future prices. Fewer studies have been made in terms of consumer decisions when faced with uncertainty. Knotek (2011) finds that households react only modestly to changes in uncertainty. I add to the literature by explicitly exploring how *demand* is affected by an increased level of uncertainty. While many of the above papers rely on a general equilibrium approach and model the whole economy, I focus on a single market. Although partial, this allows for a more specific focus on the issue.

The empirical industrial organization literature on demand estimation has focused on the issue of endogeneity of choices and qualities (see, for example, Berry, 1994; and Berry et al., 1995, for a theoretical and an empircal treament). Typically, the price elasticity is of key interest to these studies. They argue that one must take care of the endogeneity due to differences in, for example, qualities that are not observable to the econometrician, such as customer service. Similar to this paper, Berry and Jia (2010) examine the demand market for flight tickets. They deal with the endogeniety in prices and number of departures, and find that price elasticity has increased in the post-2006 market compared with the market in the period 1999–2006. The authors make this split due to large structural changes on the supply side of the US flight market occurring between the periods. Bilotkach et al. (2012) identify the price elasticity with respect to ease of access to the airport, utilizing a natural experiment in the San Francisco area in the US, namely a collapse in the highway to the Oakland, CA airport, which made it relatively less attractive compared to other airports in the area. Interestingly, Bilotkach et al. show how the distribution of price differences between San Francisco airports also shifts temporarily as a consequence of the reduced accessibility.

More specific studies about pricing behavior and demand behavior are provided by respectively Escobari (2012) and Escobari (2014) who uses a panel dataset of consumer demand and posting prices.

Using detailed 1995 data Ishii et al. (2009) argue that the product is a combination of carrier and airport contrasting Berry and Jia (2010), who argue that the product is defined by the departure-destination pair. As with Berry and Jia (2010) they define two types of consumers, namely "business" and "casual". Further, Ishii et al. link timeliness to markup and argue that consumers exhibit strong aversion to delays. The model proposed below allows for multiple types, but the estimation framework is limited to a single type of consumer.

A major complication in understanding the air travel market is its dynamic aspect of supply. Traditionally, industrial organization has emphasized the necessity of structural estimation in understanding such aspects (Nevo and Whinston, 2010).

As Berry and Jia (2010) argue, the airline supply decision is a three-step game: (i) whether to enter the market or not; (ii) choosing the supply structure, e.g. the hub, and (iii) choosing prices and frequencies and other characteristics of the offered products. The literature has mainly focused on (iii), although Berry et al. (1996) examine the endogeniety of choice of hub-and-spoke networks empirically and theoretically. Aguirregabiria and Ho (2012) examine the market structure further in a dynamic game. This paper leverages upon the existing theoretical foundations of discrete choice.

A final strand of literature relevant to this paper relates to the numerical estimation of this class of demand models. The finite properties of the Berry et al. (1995) nested fixed point iteration (NFXP) estimator have received much attention. Dubé et al. (2012) point out that the

BLP-NFXP typically used to fit the parameters and the intercepts separately, may be biased if the constants (estimated in the inner loop) are estimated with a low threshold [1], and propose a constrained optimization instead, denoted *mathematical program with equilibrium constraints* (MPEC). Reynaerts et al. (2012) examine alternative iteration procedures to NFXP and MPEC and advocate the Roland et al. (2007) *Squared extrapolation methods for accelerating fixed-point iterations* (SQUAREM) (see also Varadhan, 2012). The latter method is more intuitive, following the tradition of Rust (1987). Lastly, Skrainka (2012, chapter 2–3) proposes alternatives to the traditional simulation-based approach used in the literature, e.g. Train (2009). This paper uses a standard approach to the estimation, following Berry et al. (1995).

2.3. A model of demand for flights

Below, I discuss a theoretical model of the demand for flights with emphasis on the demand side. On the demand side I introduce a standard discrete choice framework, where agents act as if optimizing some objective function (McFadden, 1981; Train, 2009). I show how macroeconomic variables can be introduced as exogenous variables in this framework. Section 2.3.1 introduces the demand side of the market. I sketch the supply side as an oligopolistic market competing in a dynamic Bertrand-Nash game: that is each firm sets its price to maximize profits conditional on a downward-sloping demand curve. However, the supply side is not at the core of the model and serves only to introduce potential instrumental variables. Throughout, I discuss empirical and econometric implications of the proposed data generating process.

2.3.1. Demand

I first discuss the behavioral assumptions of agents and then show how this aggregates to market shares.

Behavioral model.

This section outlines the behavioral model of demand for air travel. The main assumption is that decision-makers act rationally to the extent that they maximize an objective (utility) function, satisfying the axiom of rationality (Mas-Colell et al., 1995). I proceed from a general specification concerning discrete choices to a more specific formulation aiding the empirical estimation in section 2.5.

1 This may also induce bias. A better and more Bayesian, strategy would be to utilize the fact that these markets are such that it is unprofitable for incumbents to enter.

An agent can consume at most one travel good at a given time. Physically, a decision-maker can at most board one airplane at any given time, which makes the assumption reasonable. It makes the estimation simpler, as decision-makers consume their most preferred good which may be not to choose air travel. In multi-demand models order statistics must be considered.

As such, decision-makers, indexed by n, choose a travel product from the set of feasible products T. Since flight transport is technologically constrained to departing from and landing in airports T must necessarily have a finite cardinality. Thus, estimating a discrete choice model is feasible. Not all products may be available in every market. A consumer n belongs to a market m and chooses among a set of market-specific goods, $T_m \subset T$. For instance, the combination of airports and destinations may define a product. I use the origin and the destination below. Letting U_{nmj} denote the utility (the objective function) for consumer n in market m consuming good j, we denote the consumers choice i_n. Without loss of generality this choice must satisfy

$$i_n = \left\{ i : \; i \in T_m \wedge U_{nmi} > U_{nmj} \, \forall j \in T_m \setminus i \right\} \iff i_n = \arg\max_j (U_{nmj})_{j \in T_m}.$$

While there could be situations where $\exists \imath, \jmath \in T_m : \; U_{nm\imath} = U_{nm\jmath}$ this option is not considered explicitly. It is unlikely, and if the situation occur, I assume the decision maker chooses one among the alternatives yielding equivalent utility with equal probability.

The qualities, which are variables in the objective function, affect the choice i_n. Following Dubé et al. (2012), I characterize a good j by the vector $(x_{mj}, p_{mj}, \xi_{mj})$ where x_{mj} is a vector of observable good characteristics, such as route distance, number of stops, carrier dummies, etc. and exogenous market characteristics, p_{mj} is the price of the good and ξ_{mj} characterizes unobservable characteristics, e.g. the quality or comfort of the gate area in the particular market. That is, ξ_{mj} is not observable to the econometrician but the consumer observes the full triplet. Thus, I partition U_{nmj} into a part of variables that are observable to the econometrician and a part of variables that are unobservable to the econometrician:

$$U_{nmj} = V_{nmj}(p_{mj}, x_{mj}; \beta_{nm}) + \epsilon^*_{nmj} \quad \text{s.t.} \quad \epsilon^*_{nmj} \coloneqq \epsilon_{nmj} + \xi_{mj}. \tag{2.1}$$

V_{nmj} is the expected utility given *observable* characteristics while ϵ^*_{nmj} contains the unobservable part (Train, 2009). The starred "error" depends on the characteristics unavailable to the econometrician as $\epsilon^*_{nmj} \equiv U_{nmj} - V_{nmj}$. ϵ_{nmj} is IID type I extreme value. The expected part of utility depends on a consumer-specific coefficient vector β_{nm}, where the subscripts denote the assumption of heterogeneous agents, capturing horizontally differentiated preferences.

For convenience linear utility is usually assumed (e.g. Berry et al., 1995; Dubé et al., 2012). Under linear utility (2.1) may be reexpressed as

$$U_{nmj} = \alpha_{nm} p_{mj} + x'_{mj} \beta_{nm} + \xi_{mj} + \epsilon_{nmj}. \tag{2.2}$$

Using the assumption of the random coefficients similar to Berry et al. (1995)

$$\begin{pmatrix}\alpha_{nm}\\ \beta_{nm}\end{pmatrix} = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} + \Sigma\nu \quad \text{s.t.} \quad \nu \overset{iid}{\sim} \mathcal{N}. \tag{2.3}$$

where Σ is obtained via Cholesky decomposition and captures the correlation between preferences. Both Berry et al. (1995) and Dubé et al. (2012) take Σ to be a diagonal matrix, but one need not impose this restriction. For instance, Nevo (2001) also estimates off-diagonal elements. For now, denote the parameters to be estimated in (2.3) as $\theta = (\theta_1, \theta_2)$ where $\theta_1 = (\alpha, \beta)$ and $\theta_2 = \operatorname{vec}\Sigma$. With equation (2.3) the utility specification in equation (2.2) can be rewritten as

$$\begin{aligned} U_{imj} &= \left(\alpha p_{jm} + x'_{jm}\beta + \xi_{jm}\right) + \left[p_{jn}\ x'_{jn}\right]\Sigma\nu_{nj} + \epsilon_{njm} \\ &=: \delta_{jm} + \mu_{nj} + \epsilon_{nmj}. \end{aligned} \tag{2.4}$$

δ_{jm} represents the mean utility of product $j \in T_m$ across consumers and is therefore a constant for a given j, m pair. μ_{nj} across over consumers according to some distribution and depends on the particularities of Σ of ν. For instance, if $\Sigma = \mathbf{0}$ (as assumed in the estimation) then the only random component in equation (2.4) is ϵ_{njm} which is IID by construction.

Macroeconomic variables affect demand though the exogenous variables, x_{mj}. For instance, theory would suggest that if the future is uncertain, some investments may be delayed (say business travel) and a similar argument could be made for the consumption of goods (to the extent that credit markets are not perfect). In a more complete model, one should model this as a dynamic decision, however that is beyond the scope of this paper.

The aggregated market shares.

With the basic behavioral model above we can construct the market shares of each firm.

Asymptotically the market share of product j is equal to the average probability of choosing a given product across markets. Following Train (2009) the probability of choosing j is equal to

$$\begin{aligned} P_{jmn} &\coloneqq \Pr\left(V_{nmj} + \epsilon^*_{nmj} > V_{nmk} + \epsilon^*_{nmk}\ \forall j \in T_m \setminus j\right) \\ &= \Pr\left(\epsilon^*_{nmk} - \epsilon^*_{nmj} < V_{nmj} - V_{nmk}\ \forall j \in T_m \setminus i\right) \\ \implies s_{jm} &= \int_\epsilon \mathbb{1}\left\{\epsilon^*_{nmj} - \epsilon^*_{in} < V_{nmj} - V_{nmk}\ \forall j \in T_m \setminus i\right\} dF_\epsilon(\epsilon) \end{aligned}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Assuming $\epsilon_{nmj} \overset{iid}{\sim} \mathcal{EV}$ and using (2.4)

$$s_{jm} \mid \theta = \frac{\exp(\delta_{jm} + \mu_{nmj}(\nu))}{\sum_k \exp(\delta_{km} + \mu_{nmj}(\nu))}$$

and consequently

$$s_{jm} = \int_{\nu} \frac{\exp(\delta_{jm} + \mu_{nmj}(\nu))}{\sum_k \exp(\delta_{km} + \mu_{nmk}(\nu))} dF(\nu). \tag{2.5}$$

When the model contains an outside good with a utility of zero the denominator simplifies to $1 + \sum_{k'} \exp(\delta_{k'm} + \mu_{nmk'})$.

The main difficulty lies in "inverting" or linearizing equation (2.5).

A special case where equation (2.5) can be inverted analytically is estimated below. In the general case where ν is random the BLP estimator of Berry et al. (1995) is used for estimation.

Economically, equation (2.5) is rather intuitive. A decision-maker maximizes his utility across the product space in his particular market. In particular, the decision maker may find it optimal not to participate in the market, in which case we say that he consumes the outside good with a normalized utility of zero. The market share of good j is the percentage of people who find it optimal to consume product j. The econometrician sets the parameters to make fitted market shares equal to the observed market shares. The expression utilizes the asymptotic isomorphism between market shares and average aggregated choice probability. In general, this equality need not hold in "small" samples, which in turn can affect the properties of the estimator.

2.3.2. Firms

The following is the supply side of the model. Essentially, firms are competing in a dynamic three-stage game. Let J_m be the set of firms in a market m as above. We assume firms are profit maximizing. Firm i makes the following decisions.

1. Decide whether to enter or leave the market. If firms have fixed costs this is a non-trivial decision (Hopenhayn and Rogerson, 1993).

2. Decide on product offerings, in particular airport hubs, conditional on the expectation of the product portfolios of competitors $J \setminus i$.

3. Decide product prices conditional on the realization of the product space.

This implies that prices can be changed "faster" than offerings. This, in turn, implies that the third round is a Bertrand-Nash competition yielding profits of zero or more to competitors.

The profit in a given period is given by

$$\pi(p_a; \Xi) = \sum_{m \in M} - F_{am} + \sum_{j \in T_{am}} s_{mj}(p_{amj}, p_{-mj}, \Xi)(p_{amj} - mc_{amj}) \tag{2.6}$$

where mc_{amj} is the marginal cost for a given market. F_{am} represents the fixed costs associated with operating in a given market, for instance airport fees. Ξ summarizes the state of the economy, e.g. the product space. Then, $p_a^* = \arg\max_p \pi(p_a; \Xi)$ is the price set by a given airline conditional on the current state of the economy.

From an econometric perspective (2.6) provides a basis for orthogonal moments, such as so-called cost shifters. These are most easily seen by letting mc_{amj} be linear, e.g. $mc_{ajm} = \tau_a' w_{ajm} + u_{ajm}$. Then w_{ajm} represents cost shifters. These are orthogonal since the price can be described as the markup plus marginal costs. Rival characteristics also affect demand, and it seems reasonable to assume that these are set independently.

While the BLP framework allows for joint estimation of the supply side and the demand side, I only have demand side data for the purpose of this paper.

2.4. Data

I define a *market* as a directional connection between two cities as in the literature. I choose cities rather than airports because of cross-substitutions. In the literature, one can both find markets defined as both pairs of cities and pairs of airports.

I use the DB1B dataset provided by the Research and Innovative Technology Administration (2013) using the data for 2002Q1-2013Q4 (both included). The data is quarterly and contains a sample of 10% of issued tickets in the US, including price and route details. While the data may have legs abroad, I limit the application to domestic travel within the US. I further restrain the analysis to the 50 biggest cities and 57 biggest airports in the US, as examined by Aguirregabiria and Ho (2012). The full list of airports is shown in Table 2.1. Monopoly markets (recall that a market is a city-pair) are omitted, as I cannot calculate competitor characteristics for these markets.

I further restrict to round trips as is typically done in the literature. I only consider trips costing more than USD 50. I drop flights with more than two stops, and trips that are longer than 4,500 miles[2].

This covers about half of the observations in the dataset and leaves me with approximately 2,500,000 observations.

2 I have tested several models, and estimated variables are generally volatile when including too much of the tails.

Table 2.1.: List of included airports.

City	Airport	City	Airport	City	Airport
New York, NY	LGA	San Francisco, CA	SFO	New Orleans, LA	MSY
New York, NY	JFK	Columbus, OH	CMH	Cleveland, OH	CLE
Newark, NJ	EWR	Austin, TX	AUS	Sacramento, CA	SMF
Los Angeles, CA	LAX	Memphis, TN	MEM	Kansas City, MO	MCI
Burbank, CA	BUR	Minneapolis, MN	MSP	Atlanta, GA	ATL
Chicago, IL	ORD	Baltimore, MD	BWI	Omaha, NE	OMA
Chicago, IL	MDW	Charlotte, NC	CLT	Oakland, CA	OAK
Dallas, TX	DAL	El Paso, TX	ELP	Tulsa, OK	TUL
Dallas-fort Worth, TX	DFW	Milwaukee, WI	MKE	Miami, FL	MIA
Phoenix, AZ	PHX	Seattle, WA	SEA	Colorado Springs, CO	COS
Houston, TX	HOU	Boston, MA	BOS	Wichita, KS	ICT
Houston, TX	IAH	Louisville, KY	SDF	St Louis, MO	STL
Houston, TX	EFD	Washington, DC	DCA	Santa Ana, CA	SNA
Philadelphia, PA	PHL	Washington, DC	IAD	Raleigh/Durham, NC	RDU
San Diego, CA	SAN	Nashville, TN	BNA	Pittsburgh, PA	PIT
San Antonio, TX	SAT	Las Vegas, NV	LAS	Tampa, FL	TPA
San Jose, CA	SJC	Portland, OR	PDX	Covington, KY	CVG
Detroit, MI	DTW	Oklahoma City, OK	OKC	Ontario, CA	ONT
Denver, CO	DEN	Tucson, AZ	TUS	Buffalo, NY	BUF
Indianapolis, IN	IND	Albuquerque, NM	ABQ	Lexington, KY	LEX
Jacksonville, FL	JAX	Long Beach, CA	LGB	Norfolk, VA	ORF

I combine this data with the on-time performance records also provided by the Research and Innovative Technology Administration, and a number of local and nation-wide characteristics. In particular, for each market, I find the *yearly* change of average GDP per capita, which reflects the economic state of the city (Bureau of Economic Analysis, 2015), as well as the US nation-wide GDP growth, and the change in uncertainty, as measure by Baker et al. (2015)[3].

The airports and the connection, disregarding intermediate stops, are shown in Figure 2.1. As the figure shows, densely populated areas of the US, such as the coasts, are well represented, whereas some states, such as Minnesota, are not represented at all.

I am interested in aggregated market shares, which must be endogenously generated for each good. Each good is described by a triplet $(p_{mj}, x_{mj}, \xi_{mj})$, where p_{mj} is the price of product j in market m, x_{mj} represents observed characteristics of the product, and ξ_{mj} represents

3 I am not aware of local measures of uncertainty, though it may matter. Certain areas, e.g. Detroit before the crash of its car industry, would undoubtedly face higher uncertainty than areas with a more stable industry.

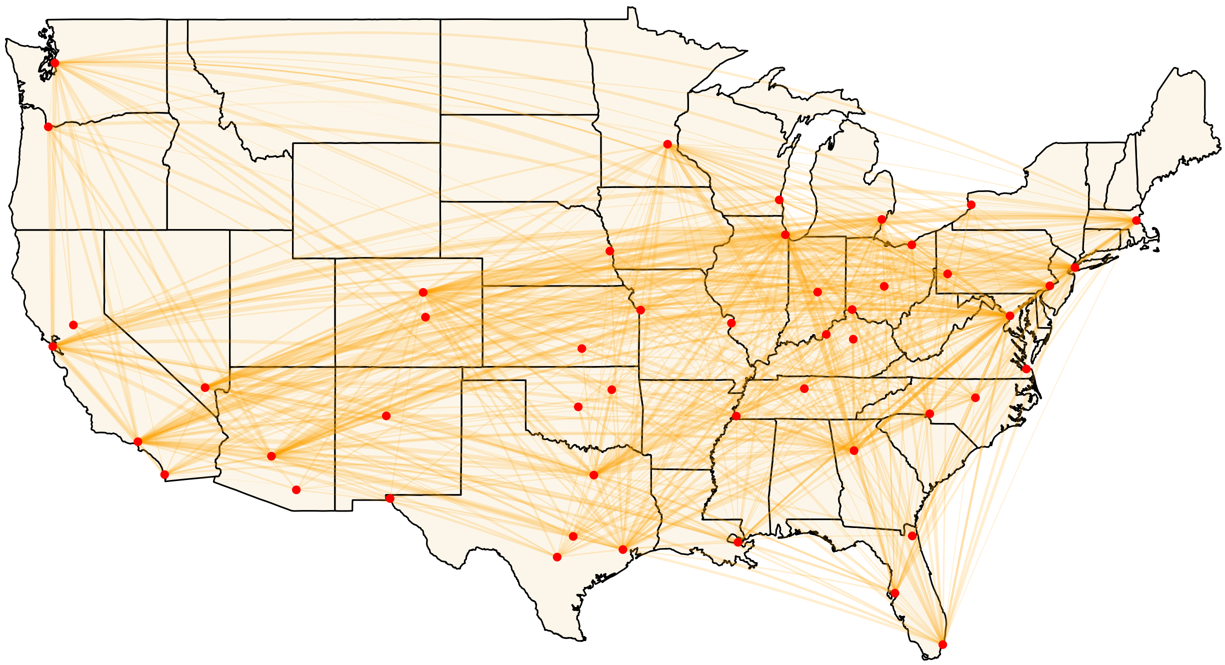


Figure 2.1.: Map of the considered airports and the connections from the data. Intermediate stops are not included in the graph, and connections are summed up. The thickness of the lines indicate how interconnected two nodes are.

unobserved characteristics. The econometrician observes measures of p_{mj} and x_{mj} and the initial task is to construct p_{mj} and x_{mj}. This is a somewhat delicate task between having too many goods and losing too much information. In particular, note that the product $(p_{mj}, x_{mj}, \xi_{mj}) \neq (p_{mj} + \epsilon, x_{mj}, \xi_{mj})\ \forall \epsilon \neq 0$, even for small ϵ. Therefore, for a market there may exist countably infinite many goods. While the discrete choice theory works with $\aleph_0$ sets, linear algebra and computers do not. Therefore, the product space must be reduced. I do this both by means of binning and by means of omission. In the aggregations of the observations, I bin characteristics such as prices and miles to lower the dimensions of the dataset.

Binning is not a neutral transformation, as the endogenously generated goods no longer reflect what consumers observed in their decision-making. However, the assumption is that there exists an economically meaningful ϵ such that consumers are (almost) indifferent between goods for which the distance between them is ϵ.

An example of a good is,

```
% (Albuquerque → Chicago, 1200 miles, $100, American Airlines, 0 stops)
```

which is the route (market) Albuquerque to Chicago, with a total distance of 1200 miles, a price of USD 100, operated by American Airlines and with zero stops. Both tickets sold at USD 95 and at USD 110 would contribute to the share of this good.

Note that the task is potentially harmful, as I essentially remove information about product characteristics. As is known from the linear case, the parameter estimated on a variable with measurement error goes towards zero[4].

In the subsequent subsections I examine each aspect of the utilized data.

2.4.1. The characteristics of the products

Each product is defined by the following observable qualities, x_{mj}: (i) distance; (ii) carrier; (iii) realized delay (measured over different types of delay); (iv) number of stops, and (v) absolute or relative number of connections from airport. In addition, the departure and destination airports are observed.

These measures have been used in the literature. However, the set of characteristics used here is more parsimonious than what is typically found in the literature.

Distance measures nuisance associated with traveling over a great distances. The carrier measures specific fixed effects associated with a given carrier. For instance, Delta Airlines may be associated with a higher quality than, e.g. JetBlue. There may, however, also be regional differences in how well a particular carrier performs. The number of stops measures the nuisance associated with a particular route.

Many airlines operate with central hubs connecting different locations. The number of connections from an airport measures how "engaged" a given carrier is in a certain airport. (iv) and (v) are generated from the DB1B data whereas the two other characteristics are directly observable. In addition, the prices are observed.

Table 2.2 displays summary data for product summary statistics across years for each of the characteristics. As the table show, the basic product characteristics are similar across the years, although delays were lower in the earlier years. The mean distance is between 2,000

4 The bias of the bivariate OLS is in this case $\operatorname{plim}\hat{\beta}_{ols} = \beta + \frac{\operatorname{Cov}(x_t^*, u_t^*)}{\operatorname{Var}'(x_t^*)} = \beta + \frac{-\beta\sigma_v^2}{\sigma_{x^*}^2+\sigma_v^2} = \beta\left[1 - \frac{\sigma_v^2}{\sigma_{x^*}^2+\sigma_v^2}\right]$.

Table 2.2.: Summary statistics on product characteristics. The plain numbers show the mean value and the numbers in parntheses are standard deviations.

Year	Stops	Distance	Connections	Delays	Carriers
2001	0.56	2408.78	28.45	7.55	11
	(0.77)	(1407.84)	(26.33)	(5.28)	
2002	0.57	2479.40	27.21	3.95	10
	(0.78)	(1429.40)	(25.38)	(3.86)	
2003	0.56	2409.89	28.49	9.23	15
	(0.77)	(1380.48)	(27.51)	(8.14)	
2004	0.55	2384.98	29.06	11.06	16
	(0.76)	(1379.97)	(28.41)	(4.41)	
2005	0.54	2303.10	29.62	12.46	18
	(0.75)	(1361.48)	(29.43)	(4.64)	
2006	0.51	2153.89	29.90	13.75	18
	(0.73)	(1321.12)	(29.83)	(5.50)	
2007	0.50	2162.34	29.71	15.03	17
	(0.72)	(1327.32)	(29.07)	(5.86)	
2008	0.48	2106.06	30.22	14.02	18
	(0.71)	(1302.14)	(28.79)	(6.97)	
2009	0.49	2129.75	28.49	11.77	18
	(0.72)	(1251.92)	(27.17)	(6.20)	
2010	0.49	2104.68	30.15	11.81	17
	(0.72)	(1223.68)	(27.72)	(5.42)	
2011	0.50	2157.85	30.17	12.50	15
	(0.72)	(1230.72)	(27.84)	(7.75)	
2012	0.47	2158.60	31.76	12.47	14
	(0.68)	(1232.29)	(28.41)	(6.80)	
2013	0.46	2126.15	32.64	13.57	15
	(0.67)	(1222.90)	(28.95)	(6.84)	

and 2,500 miles and the majority of the trips are less than 5,000 miles. The point estimate of the travel distance can also be seen to drop between 2005 and 2006, although only by a few hundred miles.

The connection variable shows how many connections a given airline has in a particular market. As the table shows, this variable fluctuates across markets, but is relatively stable over time.

Table 2.3.: Summary statistics on market characteristics and macroeconomic measures. The plain numbers show the mean value and the numbers in parentheses are standard deviations.

Year	Market Size	Local Growth	National Growth	Uncertainty
2002	4971615	-0.61	0.76	105.22
	(3200355)	(1.37)		(12.87)
2003	4895188	-1.10	1.49	110.08
	(3109287)	(1.73)		(20.67)
2004	4828057	-2.24	2.39	93.01
	(3128730)	(1.56)		(2.73)
2005	4668114	-2.36	2.29	71.85
	(3099146)	(1.62)		(4.80)
2006	4623116	-1.30	1.66	71.27
	(3144859)	(2.40)		(5.08)
2007	4591987	-0.16	0.53	80.67
	(3129249)	(2.16)		(12.72)
2008	4502918	2.17	-1.77	127.22
	(3082648)	(2.28)		(23.29)
2009	4587260	5.20	-3.87	143.70
	(3082641)	(2.86)		(18.91)
2010	4545712	-0.75	1.14	155.73
	(3061106)	(1.85)		(14.10)
2011	4752578	-0.73	0.68	172.39
	(3145480)	(1.71)		(26.82)
2012	4835504	-1.82	1.67	167.95
	(3222126)	(1.70)		(9.91)
2013	4764938	-0.78	0.84	120.09
	(3213720)	(1.12)		(7.90)

2.4.2. Market and macroeconomic characteristics

I combine the product-level data with data describing the economic conditions of each of the markets. These include the per capita output of the city and economic uncertainty. Table 2.3 display the main characteristics per year at an aggregate level.

2.4.3. Instruments

As mentioned, I expect p_{jm} and unobserved characteristics to be endogenous, and they should therefore be instrumented. Using the standard notion of Cameron and Trivedi (2005) where the data matrix is $\boldsymbol{X}$, the choice vector is $\boldsymbol{Y}$, and the instrument matrix is $\boldsymbol{Z}$, a good instrument must have two properties: (i) $\boldsymbol{X'Z} \xrightarrow{p} \Sigma_{ZX}$, and (ii) plim $\boldsymbol{Z'\epsilon} \xrightarrow{p} \boldsymbol{0}$[5].

I use two sets of instruments to identify reduced-form parameters. The first set consists of rival characteristics defined as

$$\tilde{Z} \coloneqq \left[f(\tilde{X}_{(-1)}) \ \cdots \ f(\tilde{X}_{(-i)}) \cdots f(\tilde{X}_{(-n)}) \right]$$

where f is some mapping, $\tilde{X} \subset X$ is a subset of (rival) characteristics and subscripts denote left-out observations. In the analysis, I let f be the mean of the vector and i denote market goods. The crucial assumption is that good characteristics are exogenous to the extent that they are set before prices. The market characteristics instruments employed are: (i) average or squared distance of competitors in the market, and (ii) average stops among competitors in the market both estimated before and after collapsing the dataset. Further, as in Aguirregabiria and Ho (2012) I consider: (iii) average number of direct connections of competitors in the market, to measure the hub size. The assumption allowing to be an instrument, is that hub decisions are made before price decisions.

The second set of instruments affects marginal costs and includes exogenous causes of delays, such as weather delays. If, for instance, a plane is delayed due to the weather, it may miss its departure slot. This increases the marginal costs to the provider substantially, and therefore a profit-maximizing firm would try to minimize delays (Berry and Jia, 2010).

I take macroeconomic variables to be exogenous at the consumer and provider levels.

2.4.4. Product shares

The Product shares are determined in terms of the population size of the *metropolitan* area surrounding the airport. I define the market share as the number of observations for a given (endogenously defined) good divided by the population in the area, which is the maximum market size. To make the data comparable to other studies that typically use observations for a full year, I multiply the observed number of customers by 40: four due to the fact that I want the yearly equivalent and 10 to account for the fact that it is a 10 pct. sample. The share of the outside good is defined as $1 - \sum_{j \in m} s_{mj}$.

The approach is similar to Aguirregabiria and Ho (2012), although they use the population size of the principal city of the airport, which gives a much smaller market. This assumption

5 The intuition of this can be seen directly by expanding the simple IV-estimator: plim $\hat{\beta}_{iv}$ = plim$(\boldsymbol{X'P_ZX})^{-1}\boldsymbol{X'P_zy} = \beta +$ plim $\boldsymbol{X'Z}(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'\epsilon}$. In general weak instruments bias the estimate towards the OLS estimate in the linear case. For a detailed account of weak instruments and GMM see Stock et al. (2002).

implies that people from, for example, Philadelphia will not use New York airports. Thus, I estimate the size of the market in New York to be approximately 23 million people whereas (aguirregabiria12)(nil) estimate it to be around 8 million people.

2.5. Results

Below I assume homogeneous preferences easing the computation. Further, as we are interested in the impact of an exogenous variable, the IIA condition may be of less concern in this setup. Consider the utility specification

$$u_{njm} = \alpha p_{nj} + x_{jm}\beta + \xi_{nj} + \epsilon_{njm}$$

where ϵ is distributed as a type I extreme value. The mean or expected utility is $\delta_{jn} \coloneqq \alpha p_{nj} + x_{jm}\beta + \xi_{nj}$ and choosing i implies $\delta_{im} + \epsilon_{njm} > \delta_{km} + \epsilon_{nkm} \ \forall\ k \neq i$. Integrating over the unobserved yields

$$\begin{aligned} s_{jn} &= \int \cdots \int \mathbb{1}(u_{ijn} > u_{ikn} \ \forall\ j \neq k) \Pr(\epsilon_{1n}, \dots, \epsilon_{iJn}) \\ s_{jn} &= \frac{\exp(\delta_{jn})}{1 + \sum_k \exp(\delta_{kn})}. \end{aligned}$$

where the one in the denominator follows from the implicit assumption of an outside good yielding zero utility ($\exp(0) = 1$). More explicitly,

$$s_{0m} = \frac{1}{1 + \sum_k \exp(\delta_{kn})}.$$

This can be used to linearize the shares and hence the choice probabilities:

$$\begin{aligned} \frac{s_{jm}}{s_{0m}} &= \frac{\exp(\delta_{jm})}{1 + \sum_k \exp(\delta_{km})} \left(\frac{1 + \sum_k \exp(\delta_{km})}{1} \right) \\ &= \exp(\delta_{jm}) \end{aligned}$$

taking log

$$\ln\left(\frac{s_{jm}}{s_{0m}}\right) = \delta_{jn} = \alpha p_{nj} + x_{jm}\beta + \xi_{nj} + \epsilon_{njm}$$

This can readily be estimated with OLS. The results are shown in Table 2.4.

Table 2.4 shows the model estimated under the assumption of $\Sigma = \mathbf{0}$ cf. (2.3). This implies homogeneous consumer taste. This assumption is valid if, e.g., goods are only horizontally differentiated and consumers value all characteristics equally. While this is not a correct assumption, it is perhaps satisfactory here, as we are primarily interested in the impact of

	(1)	(2)	(3)	(4)	(5)	(6)
P	-0.00063^{***}	-0.00062^{***}	-0.00061^{***}	-0.00062^{***}	-0.00061^{***}	-0.00060^{***}
	(0.00001)	(0.00001)	(0.00001)	(0.000005)	(0.000005)	(0.000004)
1 Stop	-0.86431^{***}	-0.86506^{***}	-0.86036^{***}	-0.86595^{***}	-0.86022^{***}	-0.86202^{***}
	(0.00461)	(0.00443)	(0.00433)	(0.00442)	(0.00433)	(0.00362)
2 Stops	-1.21834^{***}	-1.20897^{***}	-1.19854^{***}	-1.20763^{***}	-1.19867^{***}	-1.19165^{***}
	(0.00586)	(0.00567)	(0.00556)	(0.00563)	(0.00555)	(0.00419)
3 Stops	-1.53729^{***}	-1.55628^{***}	-1.52990^{***}	-1.55418^{***}	-1.53033^{***}	-1.50487^{***}
	(0.01103)	(0.01088)	(0.01107)	(0.01084)	(0.01106)	(0.00863)
No. Connections		-0.00854^{***}	-0.00804^{***}	-0.00857^{***}	-0.00804^{***}	0.00154^{***}
		(0.00020)	(0.00019)	(0.00020)	(0.00019)	(0.00008)
log(local GDP)			-0.96329^{***}		-0.96591^{***}	0.42940^{***}
			(0.02846)		(0.02855)	(0.06074)
log(national GDP)			1.70154^{***}		1.63966^{***}	1.37764^{***}
			(0.21933)		(0.22162)	(0.10727)
Uncertainty				-0.00064^{***}	-0.00034^{*}	-0.00020^{***}
				(0.00004)	(0.00018)	(0.00005)
Distance dummies	*X*	*X*	*X*	*X*	*X*	*X*
Carrier dummies	*X*	*X*	*X*	*X*	*X*	*X*
City dummies						*X*
Year/Quarter dummies			*X*			*X*
Observations	4,301,768	4,301,768	4,301,768	4,301,768	4,301,768	4,301,768
R^2	0.191	0.216	0.230	0.217	0.230	0.409
Adjusted R^2	0.191	0.216	0.230	0.217	0.230	0.409

Notes: Errors are clustered on market level.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

aggregated variables on consumer choice. Technically, the logit model is correct if the IIA assumption holds. That is if the elasticities are the same among all consumers. With e.g. business travelers and tourists this may not hold. Nonetheless, the estimations in Table 2.4 yield plausible results. As discussed in the theoretical part one may worry about endogeniety here in e.g. prices. However, for the purpose of this study, we ignore this issue. In regressions, not included here, the results of interest do not change much whether endogenous variables are instrumented or not.

In total six models presented going from rather sparse to rather complex. In model (1) I control only for prices, airlines number of stops (as dummies) and distances (also as dummies) to capture potential non-linearities. As can be seen in this model, signs are as expected. Price has a negative impact on the propensity to choose a product, as does more stops. Likewise, I included dummies for carriers to capture carrier fixed effects.

Model (2) further adds number of connections in the airport for a given firm. As can be seen this has an unexpected sign suggesting further missing controls, which subsequently added in model (6).

Models (3), (4), (5) and (6) are the models of main interest. In model (3) I add the local and nation-wide economic factors and their impact on the propensity to buy. As can be seen, the sign of the local GDP effect is surprising. In model (4) I add uncertainty, which in general has the expected sign in the sense that an increased level of uncertainty leads to an increased propensity to demand goods viz. a. viz the outside good. In Model (5) I show both uncertainty and economic effects, and although local GDP has the wrong sign the remaining signs look as expected. In model six I add controls for year and quarter effects as well. As can be seen this is not unexpectedly important to make the model work.

In general model (6) does fairly well. All coefficients have the expected signs.

We next turn to the economic magnitude of the model. Since the demand model is non-linear we must compute the elasticities in order to interpret the coefficient. Table 2.5 shows the implied elasticity and semi-elasticiy for price and uncertainty. It is worth noting that the sign is as expected. The main contribution of this paper is shown in the second column of table 2.5. The result shows that uncertainty has a small economic impact on the propensity to demand goods among consumers. The elasticity of uncertainty on demand is relatively small compared the price elasticity. The price elasticity implied by the model is generally in line with the literature, e.g. Berry and Jia (2010), though the results may be slightly biased upwards.

Table 2.5.: Implied elasticities based on the model.

	Own-price	Uncertainty
Model 1	-0.359	
Model 2	-0.351	
Model 3	-0.344	
Model 4	-0.352	-0.077
Model 5	-0.344	-0.040
Model 6	-0.343	-0.023

2.6. Conclusion

In this paper I have investigated the impact of uncertainty on demand in an important market, namely the market for air transport, using a long series of US microdata. I show that throughout the period from 2002–2013, which is a period with changes in overall uncertainty and impact on the demand of consumers. This indicates that uncertainty shocks do not only affect the supply side, but also the demand side.

Chapter 3.

`forecast.js`: a module for measuring expectation in economic experiments

The first version of this paper was authored by Rasmus Pank Roulund. A subsequent version is co-authored with Nicolás Aragón (Universidad Carlos III de Madrid)

Abstract

This paper presents a elicitation tool for economic experiments based on Harrison et al. (2017). The tool is user-friendly and enables subjects to forecast the movements of a continuous or discrete variable while addressing the main problems in eliciting beliefs. The module is easy to integrate with HTML-based experimental software kits, such as oTree (Chen et al., 2016).

Expectations play a central in many economic phenomena, both in microeconomics and in macroeconomics. However, they are particularly important when analyzing the behavior of financial markets; i.e, the dynamics of panics, crashes and bubbles. Though important, rigorous empirical analysis is particularly challenging given that expectations are not directly observable. It is not surprising, therefore, that expectations have come to be analyzed heavily in experimental settings, where elicitation seems possible.

The literature has normally attempted to obtain point estimates of forecasts and, at most, a confidence band. Accuracy is normally incentivized by using monetary rewards. However, even in an experimental setup carefully eliciting expectations represents challenges.

First, by requesting agents to give a point estimate with a confidence band, there is often an implicit assumption that expectations follow a symmetric distribution, e.g. a uniform distribution. This leaves out the possibility that agents may have beliefs over two possible scenarios disjointly (for example, fundamental and bubble values in an asset market). Thus, a point estimate leaves valuable information out of the study and may even bias the elicited belief. Moreover, it does not provide information on second moments of expectations, i.e, the role of confidence in a forecast.

Second, if subjects need to perform more than one task and are incentivized for both, they may choose to diversify risk between the activities, thus hedging and distorting the elicited beliefs [1] (see e.g. Armantier and Treich, 2013b). For instance, if agents need to trade assets and provide forecasts, they may trade at a high value and report expectations at a low value, thus ensuring a less variable payoff. The literature has normally avoided giving large payments for expectations, as to not distort choices via hedging. However, if the objective of the study is to carefully analyze expectations, it can be argued that the reporting of beliefs needs to be incentivized with large expected payoffs.

Finally, risk perceptions may change the reported beliefs of subjects. For instance, more risk averse agents tend to report "flatter" distributions than their true beliefs (Harrison et al., 2015).

In this paper we present an open source Javascript module that enable clean elicitation of a *distribution* of beliefs of a discrete or continuous variable, using the method laid out in Harrison et al. (2017). The module is graphically appealing and user-friendly. It easily integrates with HTML-based experimental software, such as oTree (Chen et al., 2016). It has been successfully used in asset market experiments in Barcelona (Aragón and Pank Roulund, 2019), where it was verified that participants felt comfortable with the software and learned to use it quickly.

The source of this module can be found at `https://gitlab.com/pank/forecast.js`. Contributions, requests and improvements are welcome and can be submitted via the issue tracker there. The rest of the paper is organized as follows. Section 3.1 outlines the literature of elicitation of subjective beliefs as background to the design choices that have been made for the module. Section 3.2 explains how to use the `forecast.js` module in an experiment.[2] Section 3.3 show examples of the sort of data that has been collected with `forecast.js`. Section 3.3.2 analyzes the learning curve for users and the ease of use . Finally, Section 3.4 concludes. Appendix B shows a standalone HTML demo, and an online oTree can be found on the module's Gitlab page.

3.1. Background

In this section the theory and background literature that has led to the design choices of the `forecast.js` module are presented.

The principal task of the module is to elicit subjective expectations about a particular event, such as the movement of a continuous or discrete variable. A typical example would be the

1 If a financial trader has access to several market related to the same underlying asset he can hedge his bet in one market through actions in another market.

2 More detailed instructions are also available in Appendix B.

price of an asset in a future period, but the module can be applied to any other scenario, such as quantities to be produced or voter turnout.

Fundamentally, expectations are *subjective probabilities*, meaning the "degree of belief regarding the likelihood of events" (Karni, 2014). As such, the subjective probabilities in which we are interested follow the usage proposed by Savage (1954) and others before him (see Karni, 2014, for a full treament of the historical developments of subjective probabilities).

3.1.1. Elicitating Expectations in Experimental Finance

A number of experimental finance studies have elicited expectations from participants, especially through point estimates and confidence bands.

The workhorse of experimental finance, Smith, Suchanek and Williams (1988), elicit price beliefs in a market where a single type of asset is traded. In each round, the equilibrium price of the asset is calculated based on the participants' bids and asks. Participants are also asked to predict future equilibrium prices by giving a single-point forecast. Haruvy et al. (2007) set up a similar experimental asset market and elicit beliefs for all future prices with a confidence band.

Zarnowitz and Lambros (1987) note that it is important to examine the whole distribution of beliefs of a particular forecaster (a similar argument is put forward by Manski, 2004; and Engelberg et al., 2009).

The main drawback of using a point estimate is that it is not possible to determine the confidence or the uncertainty surrounding a prediction, nor to address the role of bimodal beliefs[3]. Moreover, to understand the degree of overlap of expectations and the diffusion of the forecasts among forecasters it is necessary to have information on the whole distribution.

Manski (2004) advocates for the usage and elicitation of subjective probabilities about events, rather than the revealed preference approach, e.g. the random utility model and discrete choice models (see Train, 2009, for an overview). Engelberg et al. (2009) use data from the Survey of Professional Forecasters and compare point forecasts and full distributions of forecasts given by the same individuals. While they find that the point forecasts tend to be close to those implied by their full distribution, they also find that the deviations tend to be asymmetric, thus implying a bias.

3 Bimodal beliefs are useful in situations where the equilibrium price differs from the fundamental value and seemingly follows a, e.g., inflated path. A participant may then suspect that either the bubble will burst and the price will return to the fundamental value or that the price will keep following the inflated path.

The approach of eliciting the full distribution of beliefs is becoming common in economics. Harrison and Phillips (2014) elicit the full subjective belief distribution from a number of professional forecasters, Chief Risk Officers (CROs), and compare those to the predictions of statistical models. They find that, in some cases, tail risks are perceived as lower by professional forecasters than statistical models.

Other examples of full distribution elicitation of expectations include the Fed Survey of Professional Forecasters (see e.g. Diebold et al., 1997), the ECB Survey of Professional Forecasters and the Bank of England Survey of External Forecasters (Boero et al., 2008). Engelberg et al. (2009) compare the point forecasts and full distributional forecasts of professional forecasters in the Fed Survey of Professional Forecasters, and find that while the point forecast tends to correspond to the central moment of the corresponding distribution, the differences that do arise tend to be positive. Note that the professional forecaster surveys are not directly incentivized in the same way as observed in economic experiments.

In this paper we present a module that enables the elicitation of these beliefs in experimental settings in a quick and user-friendly way. The module has been used in Aragón and Pank Roulund (2019) as an extension of the Haruvy et al. (2007) setup. In this way, the full distribution of beliefs is elicited, enabling measurement of the initial "agreement" among traders and their certainty.

3.1.2. Eliciting a Distribution of Beliefs: Theoretical Considerations

In the `forecast.js` module, participants report a probability mass function characterized by $\{r_k\}_{k=1}^{K}$ where K is the number of intervals in a discretized variable of interest. The sum over the grid points equals one, $\sum_{k=1}^{K} r_k = 1$, as it is a probability distribution. Once a participant's forecast distribution is known, a scoring rule needs to be used under the assumption that interval j contains the ex post realized value of the measured variable.

A key element to properly elicit expectations is a scoring rule. A scoring rule is a tool to map the elicited expectations into a score once the accuracy of the expectation is known. Scoring rules may be used to incentivize forecasters to reveal their "true" beliefs in addition to increasing their stakes and effort in the task. A scoring rule must provide incentives such that the participants can maximize their expected outcome by revealing their true beliefs. In addition to showing the mathematical aspect of eliciting expectations, Savage (1971) also discusses a number of issues with respect to eliciting expectations using scoring rules.

A particularly interesting scoring rule for the purpose of experimental economics is the quadratic scoring rule (QSR) (see Matheson and Winkler, 1976). As Harrison et al. (2017) show, the QSR—and any proper scoring rule for that matter—has a number of useful properties as shown below. We denote the reported beliefs as $\{r_k\}_{k=1}^K$ and the true beliefs as $\{q_k\}_{k=1}^K$ (see Harrison et al., 2017, section 3 for details).

1. A participant only reports positive expectations about a specific event, e.g. $r_i > 0$, if his or her true belief of the outcome is positive, i.e. $q_i > 0$.

2. If a risk-averse or risk-neutral participant believes two events are equally likely, say $q_i = q_j$, then the reported expectations are the equal, i.e. $r_i = r_j$.

3. If a risk-averse or a risk-loving participant report the same expectations for two events, e.g. $r_i = r_j$, then the true beliefs are also equal, $q_i = q_j$.

4. If a participant's true beliefs are represented by a symmetric distribution then the means of the reported beliefs and the true beliefs are equal. This is true for unimodal and multimodal distributions alike.

5. If the participant's reported expectations are a symmetric distribution, then the true belief distribution is also symmetric.

6. The more risk-averse a participant is, the more the reported distribution will resemble a uniform distribution with the support of the true distribution of beliefs.

In addition, Harrison et al. (2017) report that the reported distributions are "very close" to the true subjective distributions for a wide range of empirically plausible risk attitudes, provided individuals are Expected Utility Maximizers. A method of binary lotteries presented in Harrison et al. (2015) can risk-neutralize subjects enabling the elicitation of beliefs under more general preferences. The software can be easily adapted to perform this task.

The discretized version of the QSR can be expressed as follows (Harrison et al., 2017):

$$s_j(r) = \kappa\left[\alpha + \beta\left(2 \times r_j - \sum_{k=1}^{K} r_k^2\right)\right].$$

Here, κ, α and β are parameters. The scoring rule can be configured directly in the `forecast.js` module. Parameters α and β penalize spread forecasts, and κ is a scaling parameter that can be used to avoid hedging using proper calibrations. This is further discussed in Section 3.2.1.

3.2. Using the `forecast.js` module

This section shows how the `forecast.js` module is used and how it may be adapted to the needs of individual experimenters. The main purpose of the module is to enable experimenters

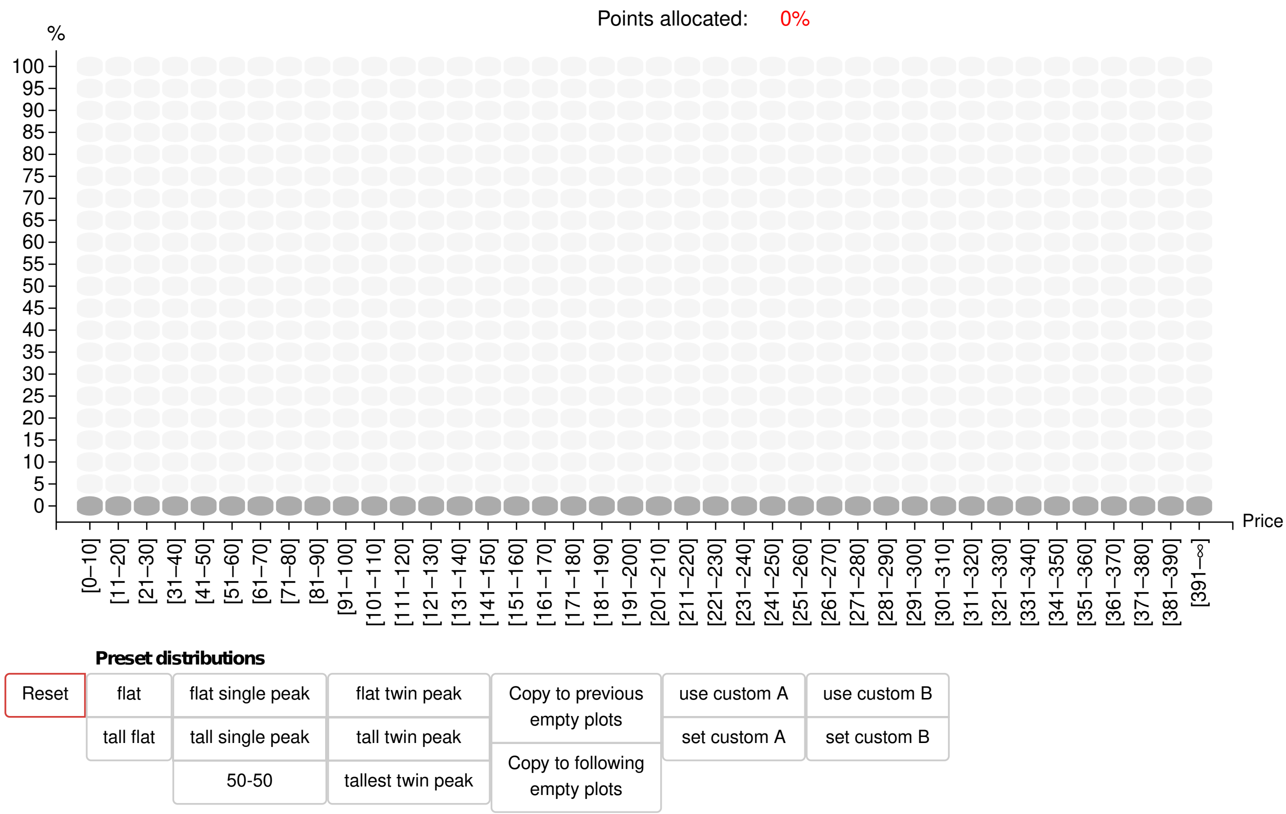


Figure 3.1.: The plain canvas presented initially to forecasters.

to easily obtain measures of a participant's beliefs about the movement of a continuous or discrete variable. By default, this is assumed to be a price.

The forecast module is written in Javascript and is therefore supported in all major browsers, such as Mozilla Firefox and Google Chrome. The `d3.js` SVG toolkit (see Bostock et al., 2011) is the primary library. As it is a simple Javascript module, it can be used with experimental toolkits utilizing web technologies, such as oTree (Chen et al., 2016). The module makes it possible for participants in economic experiments to easily produce a visual representation of their beliefs, in the form of a probability mass function, by drawing it with their mouse.

When participants are first presented with the tool they observe a "blank canvas" as shown in Figure 3.1. Along the primary axis, the value of the variable is shown. In the example, the measured variable is the price of asset shares. The domain of the canvas in the picture is $[0, 400]$. The range can be changed by setting the variables `xmin` and `xmax` before loading the module. The picture contains 40 bins, meaning that each bin covers a price range of 10. The number of bins along the primary axis is determined by the variable `xbins`, which can be set before loading the module. The bin width is automatically calculated based on `xmin`,

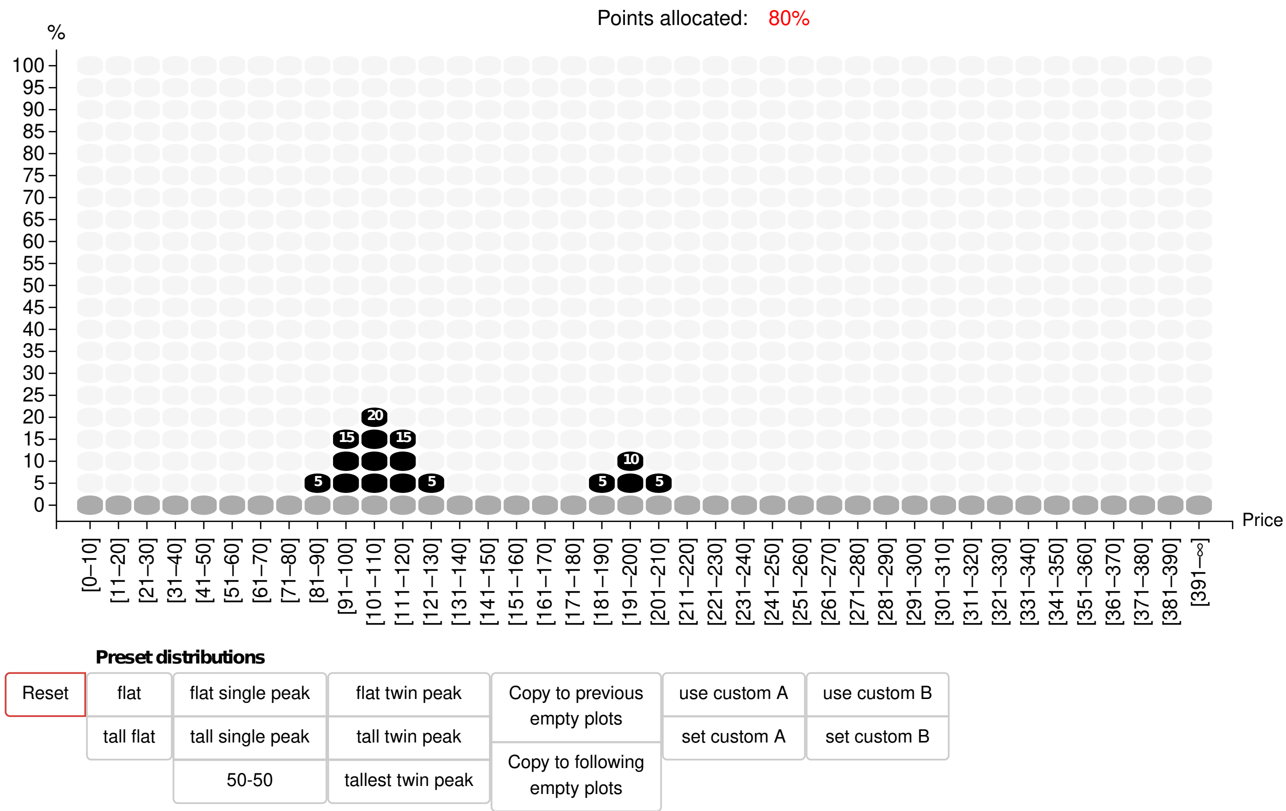


Figure 3.2.: An intermediate forecast. Note that the module warns the user that the forecast is not complete as not all "tokens" have been allocated. One difference from Harrison et al. (2017) is that, by default, no tokens are allocated as opposed to a "flat" uniform distribution. On the image shown, there are more x-bins than tokens, which is one reason not to start with a flat distribution. Scores are only calculcated once all tokens have been allocated.

`xmax` and `xbins`. By default, the name of the primary axis is "Price", but this can be changed by setting the variable `xname`.

The number of bins on the secondary axis can be controlled by setting the variable `ybins`. The default value is 20, meaning that participants have 20 tokens to allocate. In other words, each token represents 5% confidence in the bin. To make the elicitation intuitive for the participant, it is suggested that this value be kept as a multiple of 5. As the module is meant to produce a probability mass function, the range is fixed at $[0\%, 100\%]$. Nonetheless, the name of the secondary axis can be changed by setting the `yname` variable.

The size of the canvas, measured in pixel[4], can be set using the variables `fheight` and `fwidth`, representing height and width respectively.

4 Note that this refers to a so-called "css pixel", as opposed to the more familiar physical pixels. css pixels vary in size depending on the device, as explained by Chien and Nyman (2013).

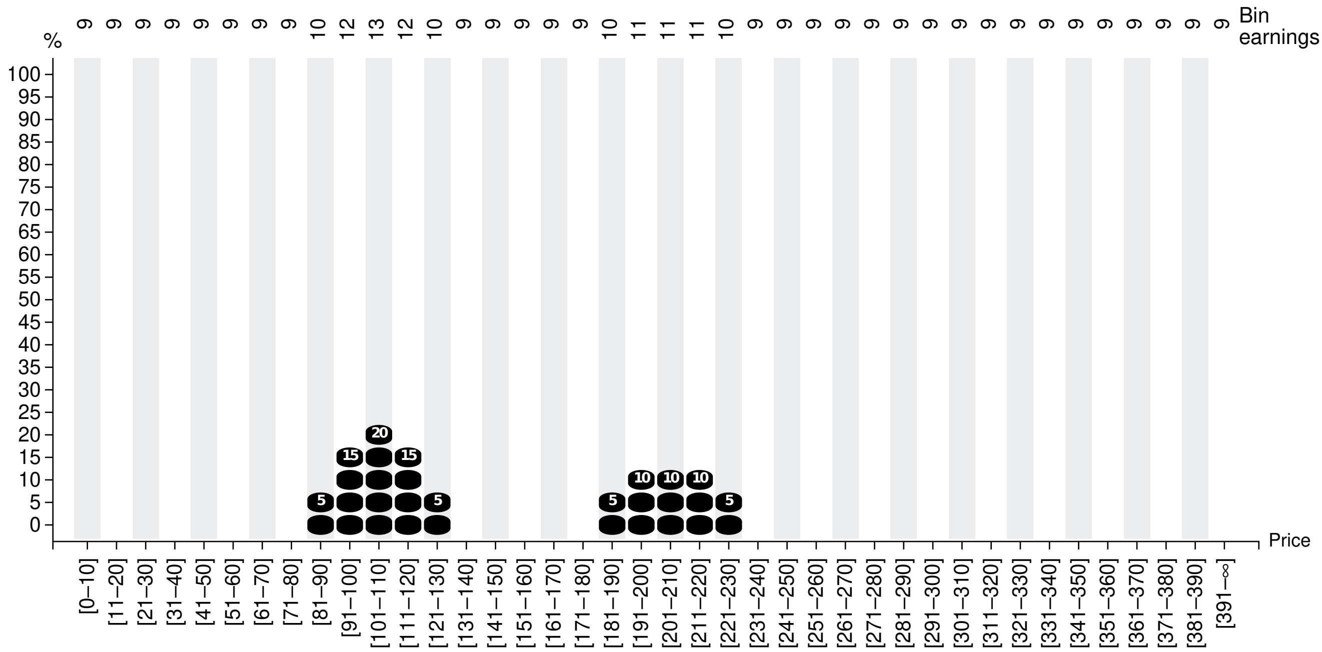


Figure 3.3.: A finalized forecast with the implied scores shown at the top of the screen.

Participants can draw probability mass functions as shown in figure 3.2 by using the mouse. The black bars denote the expectations of the forecaster. Above the tokens, the percentage value corresponding to the height of the bar in percentage terms is displayed, representing the confidence. It is possible to drag-and-drop the bars along the primary axis, or to change the height of individual bars by clicking on the tokens. The module is also endowed with customizable pre-configured probability mass functions, making it quicker for participants to forecast. It is also possible to save a specific forecast to use later. The forecast can be reset or finalized using the respective buttons.

Figure 3.3 displays a finalized forecast. A forecast can only be finalized once all tokens have been allocated by the forecaster. At the top of the canvas, the score for each bin is displayed, given the scoring rule and the forecaster's distribution. The forecaster is free to change his or her forecast after observing the scores. This can be done by clicking the "edit" button (not shown). Alternatively, the forecaster can accept the forecast and move on to the next stage of the experiment.

The score of a specific bin typically depends on the shape of the entire forecast distribution. Thus, it can only be calculated once the distribution has been finalized. By displaying the scores directly above each bin, the experimenter can focus on explaining the broad mechanism of the scoring rule (i.e, penalizing dispersion and increase in accuracy) rather than to focus on the mathematical details (this approach was also taken by Harrison et al., 2017). By default, the quadratic scoring rule is used by the `forecast.js` module, but any scoring rule can be used. To apply a different scoring rule, the experimenter can set the variable `scoring_rule` to a function accepting two arguments. The first argument is an index of the distribution to be measured (corresponding to j in the QSR, equation 3.1.2), while the second argument is the distribution. For instance, to use the logarithmic scoring rule,

$s_j(r) = \log(r_j)$ one could set,

```
scoring_rule=function(j,r){return(Math.log(r[j]))};
```

3.2.1. Calibration

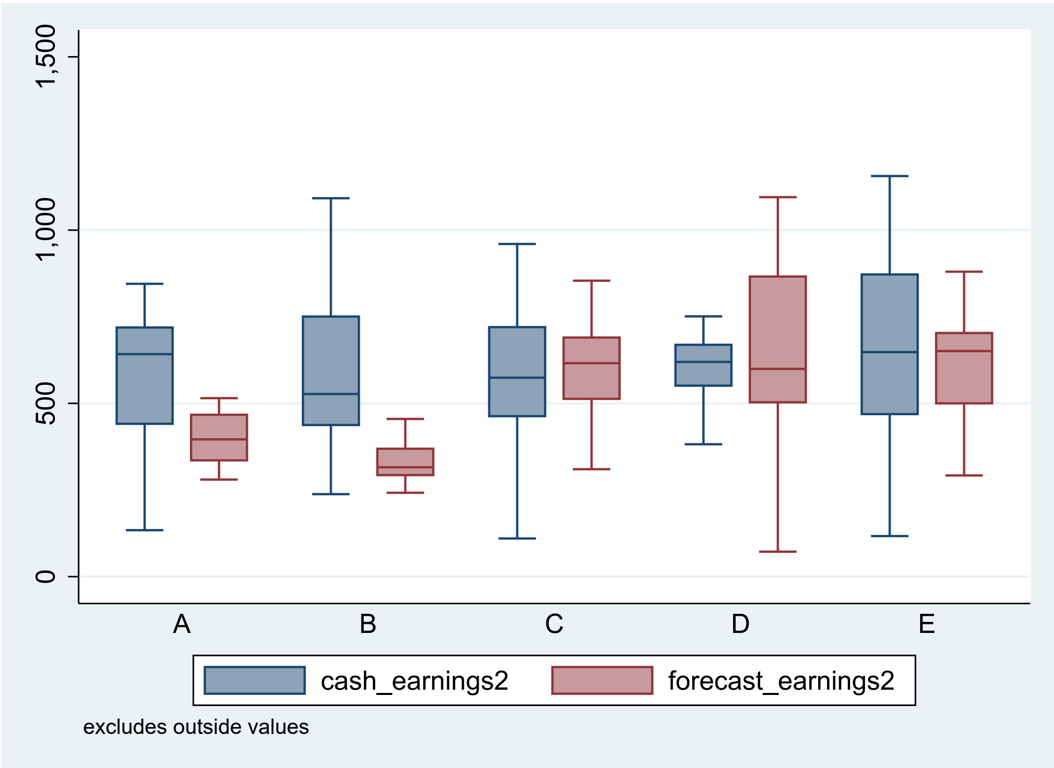


Figure 3.4.: Payment distributions across sessions. Parameters are $\alpha = 6$, $\beta = 14$ in all sessions. $\kappa = 4$ in sessions C, D and E, and $\kappa = 2$ in sessions A and B.

The parameters in the QSR can be calibrated in the module.

First, α and β must provide incentives for accurate forecasts. In Aragón and Pank Roulund (2019), α and β are chosen such that payments are either zero or slightly negative in very spread out forecasts. Irrespectively, given the repeated nature of the forecast task, final negative payments are very unlikely. In Figure 3.4, payments with $\alpha = 6$ and $\beta = 15$ are presented for both transactions and forecasting.

Secondly, the parameter κ can be used to avoid hedging in the market and to elicit expectations cleanly. Hanaki et al. (2018) show that subjects indeed show different behavior when they need to both trade and make forecasts. Thus, it is concluded that a payment based on either forecasting or trading performance chosen randomly at the end is better than paying subjects based on both. In this way, subjects need to put maximum effort into both tasks. A proper calibration of κ can rescale payments to make participants' payments ex ante equal between the two tasks. The amount of cash in the economy is fixed in expected terms, as there are the initial endowments of each participant and the expected dividends times the number of periods. Aragón and Pank Roulund (2019) use pilot sessions to calibrate average payments. Figure 3.4 shows average payments under both tasks for different parameters levels. As we can see, a value of κ of 4 under these conditions generates equal expected payments for participants under both tasks.

3.2.2. Accessing the forecast data

As the expectations about the future movement of the variable in question are elicited, the values must be collected and stored.

Typically, the experimental software will take care of storing the data once it knows how to collect it. The `forecast.js` module has two main outputs. The first one is an array representing the distribution of expectations that has been elicited. This is a standard probability distribution summing to one. It has the same number of elements as there are bins along the primary axis. The second output is a vector of scores, containing an individual score for each bin, assuming that the bin contains the correct value ex post.

To illustrate, consider an experiment where a participant forecasts some value over six bins. Consider the following example of a forecast,

$$\hat{r} = (0, 0.15, 0.35, 0.35, 0.15, 0) .$$

In this example, the participant states that the probability of the true value materializing in the interval represented by the second bin is 15%, and so forth. This value is stored as the string `"[0, 0.15, 0.35, 0.35, 0.15, 0]"` by `forecast.js`. By default, the distribution is stored in the `dist` variable. The name of the storage field can be changed by setting the variable `dist_var_prefix`. If the participant is asked to do more than one forecast, they will be stored in the variables `dist_1`, ..., `dist_n` (see Appendix for details).

Scores are also stored in a variable named, by default, `score`. The prefix can be changed by setting the variable `score_var_prefix` before loading the `forecast.js` module. Using the quadratic scoring rule with $\alpha = \beta = 10$, and using the above example distribution, the scores of $\hat{r}$ would be $(7, 10, 14, 14, 10, 7)$. These would be stored in the `score` variable as `"[7, 10, 14, 14, 10, 7]"`.

3.3. The generated data

This section illustrates the type of data that is generated with the `forecast.js` module. We first show an example of a forecast series generated with the module and then turn to the time consumption of the usage of the module. As the timing data shows, participants are generally quick at learning how to use the module and at making forecasts. Finally, we turn to the precision of module over time, to see to what extent learning takes place.

We use data from Aragón and Pank Roulund (2019). In that experiment, traders are asked to predict the movement of the asset price at the beginning of each round, i.e. a one-period-ahead forecast. The experiment follows the structure of a standard experimental market (see Smith et al., 1988). The price is determined via a call market for a single asset. The asset pays a random dividend each round, with an expected value of 12 each period. The fundamental value of the asset is determined as $f(t;T) = 12(T + 1 - t)$ where T is the total number of rounds in the market and t is the current round. In these markets, prices typically tend to start below the fundamental value, increase above it, and finally burst to a level below it (Palan, 2013).

3.3.1. Example of individual expectations

Figure 3.5 illustrates the predictions of a participant throughout an entire experiment. The vertical line displays the realized equilibrium price and the histogram illustrates the participant's forecast. The participant took part in an experiment with two consecutive markets, each consisting of 15 periods. Market 1 is shown in the left column and Market 2 is shown in the right column. Each line corresponds to a round in the market. In the example, the bin width is 10 points, corresponding to a confidence of 5%. This means there were 20 tokens to be allocated by the participant, who, in turn, was scored using a quadratic scoring rule. The graph illustrates the type of data that can be elicited with the `forecast.js` module.

In this particular example, the participant is not certain about the exact range of the price in the first period of market 1 (the left column), and thus chooses a disjoint probability mass function. However, after the second period, the participant ceases to use disjoint distributions. The participant has positive beliefs in the right price brackets for all remaining periods. In general, the participant also becomes more confident about the price development from period 1 to period 11, in the sense that the range of the distribution is overall decreasing. After market 1 crashes in period 11, the participant manages to correctly predict the price again. In market 2, the participant generally manages to correctly predict the price between period 2 and 6, and again from period 11 until the end of the market.

While this particular participant tends to spread his or her forecasts over two to three bins,

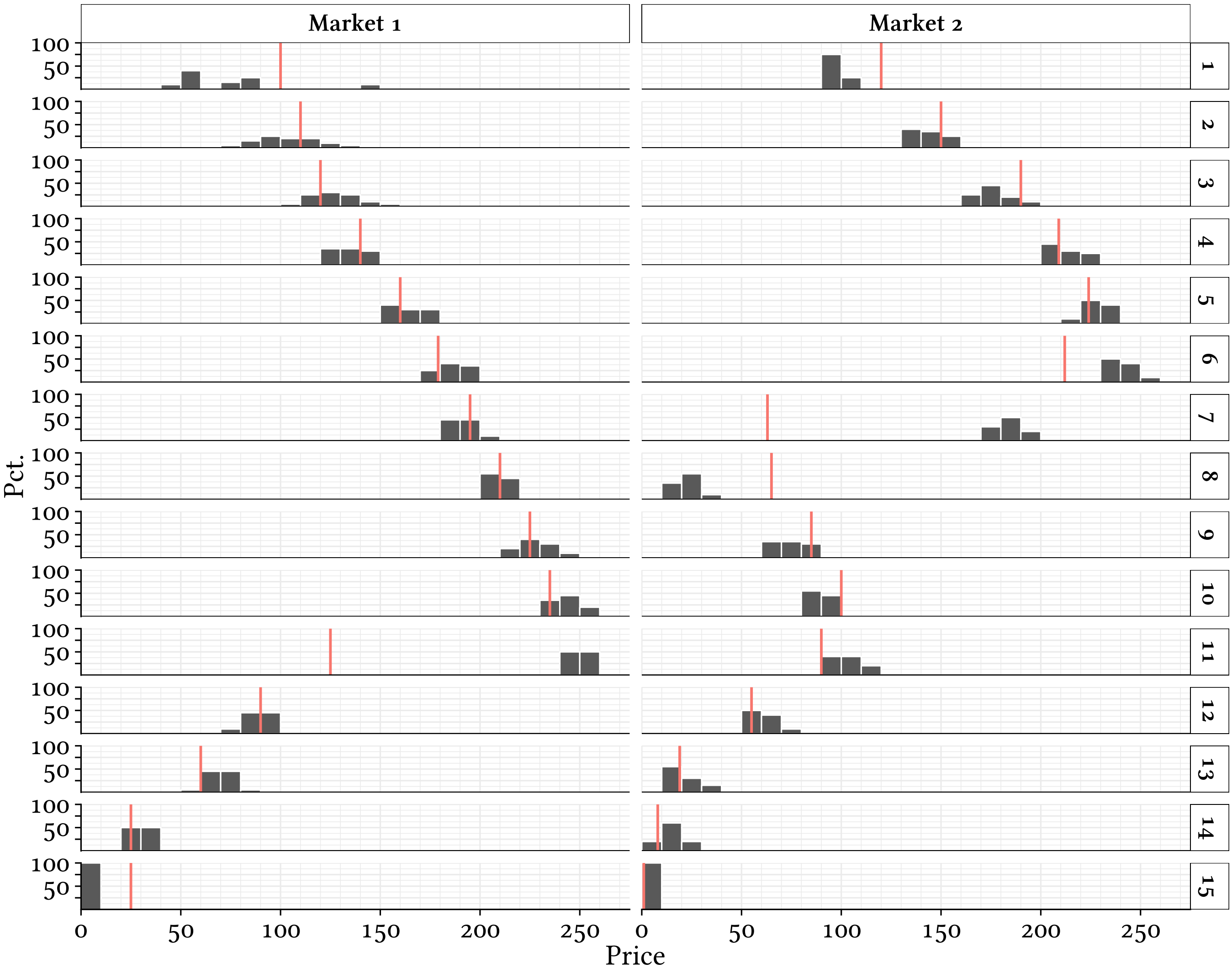


Figure 3.5.: An example of a participant's predictions. The black histogram corresponds to the price expectations of the participant, and the thin, red vertical line corresponds to the ex post realized price.

at times (s)he does use wider expectation distributions, such as in period 3 in both markets. Often, one bin, not necessarily the median, receives more weight than other bins, as shown in period 13 in market 2. As such, the example illustrates to what extent it is possible to gather more detailed information about the participant's expectations, compared with when imposing e.g. a fixed deviation of the distribution.

3.3.2. Timing Considerations

An important aspect of the `forecast.js` module is that it is quick to use for participants, so as not to delay the experiment. This means that experimenters can elicit full expectation

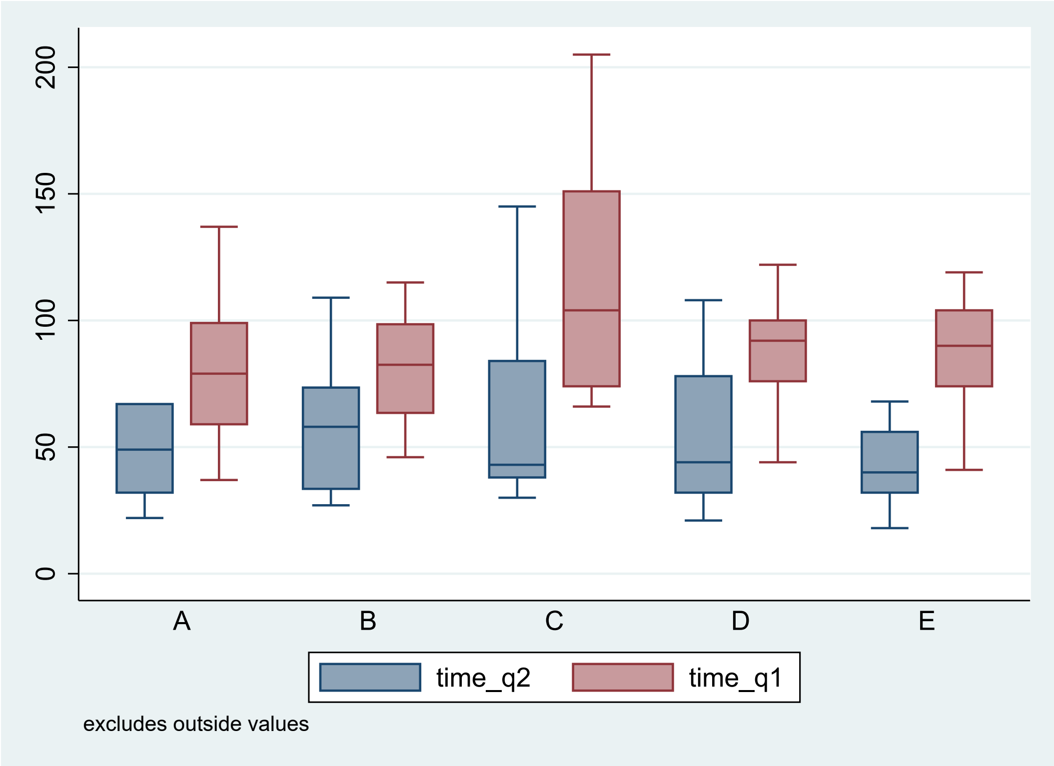


Figure 3.6.: A box plot of time spent (in seconds) on the understanding questions.

distributions from participants without a big toll on the duration of the experiment. Moreover, the results show that the elicitation time is decreasing as the participants become more accustomed to it, lowering the total time costs if an experimenter plans to elicit expectations several times.

Understanding Questions

We propose two sets of understanding questions to be used with this module as to make sure participants understand the tool. The HTML version of these questions can be found in the Gitlab page of the module, and can be easily implemented in oTree-based experiments.

The first set of questions requires the participants to get familiar with the tool. It requires them to put a certain amount of tokens in certain prices and explains the meaning. For example, they are asked to put half of the tokens in the bin corresponding to a price of 100, and it is explained that this means they think that half of the times the price will be that one. Subsequently, participants are asked to drag the created distributions, and are required to modify a preset distribution[5]. The time taken to answer these questions in the Aragón and

5 The tool used by Harrison and Phillips (2014) also used preset distributions.

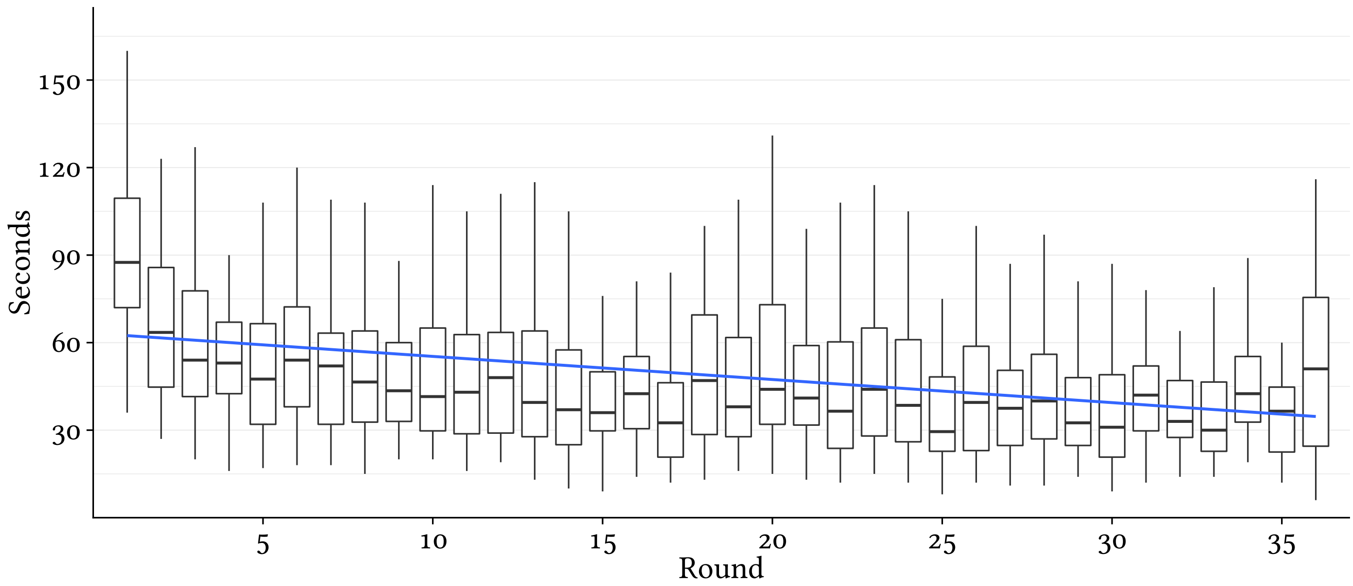


Figure 3.7.: Time spent on eliciting expectations. $N = 60$ observations per round from rounds 1–30 and $N = 24$ observations in periods 31–36.

Pank Roulund (2019) experiment is shown in Figure 3.6.

The second set of questions requires students to understand the earnings they would get once a forecast is finalized. Figure 3.6 shows the time spent on these questions. Overall, it takes few minutes for the participants to get acquainted with the tool.

Timing over time

Figure 3.7 displays a box and whisker diagram for the time spent on eliciting expectations in each round in the Aragón and Pank Roulund (2019) experiment. The lower hinge of the box corresponds to the lower quartile of the forecasting duration, the middle line is the median and the upper hinge is the upper quartile. The whiskers correspond to the largest (smallest) observation at most 1.5 times the interquartile range (i.e. the height of the box) away from the top (bottom) hinge. The line is a simple linear trend showing that, on average, participants become faster at using the module over time.

As the graph shows, most participants use between 72 and 110 seconds in the first round where they use the tool.[6] The mean and the median are 92 and 88 seconds respectively. After the first period, the median time generally decreases.

One important caveat is that we cannot disentangle to what extent the time is decreasing

6 The complexity of the situation of having to realize an unknown price without a strong prior (other than the fundamental value) may also influence the longer time we observe in the first period.

Table 3.1.: Regression showing the impact of repeated use of `forecast.js`.

	Elicitation time (seconds)			
	(I)	(II)	(III)	(IV)
(Intercept)	71.76^{***}			
	(3.05)			
t	-2.23^{***}	-2.34^{***}	-3.31^{***}	-2.72^{***}
	(0.30)	(0.29)	(0.37)	(0.38)
t^2	0.04^{***}	0.05^{***}	0.06^{***}	0.06^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Participant FE		✓	✓	✓
Session FE			✓	✓
Market FE			✓	✓
Round FE				✓
Num. obs.	1944	1944	1944	1944
R^2	0.08	0.32	0.33	0.34
Adj. R^2	0.08	0.30	0.31	0.32

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Std. errors clustered on participants.

due to increased familiarity with the module and to what extent it is decreasing because participants understand the game better.

In a regression setup we can partially control for characteristics of the underlying experimental data. Table 3.1 shows some specifications in which the time (in seconds) spent using the tool is regressed on the round, t, and a quadratic component, t^2. A set of fixed effects is included, to account for participants' characteristics and the experiment. The data contains a total of 1944 observations, split across 60 participants.

The round number corresponds to the number of times the participants have forecast the experiment. The first column (I) shows that the elicitation time decreases by approximately 2 seconds per round for the first 3 rounds, after which it drops by 1.5–1 seconds for rounds 4–15. This indicates that participants spent some time getting acquainted with the elicitation tool, but they were relatively quick to learn how to use it.

Column (II) includes individual fixed effects for each participant (the coefficients are omitted from the table as there are 60 different individuals). As can be seen, the main conclusion holds: time spent is decreasing with repeated usage of the module, but is more pronounced in the first rounds. Column (III) and Column (IV) include additional fixed effects. The results are not statistically different from the basic result in Column (I).[7]

7 In principle, more complex models could be used. In regressions not shown, we have included variables such as the lagged difference between the realized price and the fundamental value, the volatility of forecasts in

In summary, Figure 3.7 and Table 3.1 show that the elicitation of participants' expectations is fast with the `forecast.js` module, even in a relatively complex situation, and the time spent per elicitation decreases with repetition.

3.3.3. Prediction precision over time

It is important that the elicited expectations are consistent across the experiment and that the quality of the elicitation is not hindered by the complexity of the elicitation tool. In this section it is examined whether people tend to perform better as they gain experience with the elicitation tool.

Figure 3.8 displays normalized scores across sessions[8] for each period. The thick line displays

the previous round, the correction needed to align the mean/median of the past forecasts with the realized prices, etc. These variables do not have significant explanatory power and are thus not shown here.

8 Scores are normalized within experimental sessions to make them comparable across sessions. For a given session, normalized scores, $\hat{x}$, are calculated based on the "plain" scores, x, as

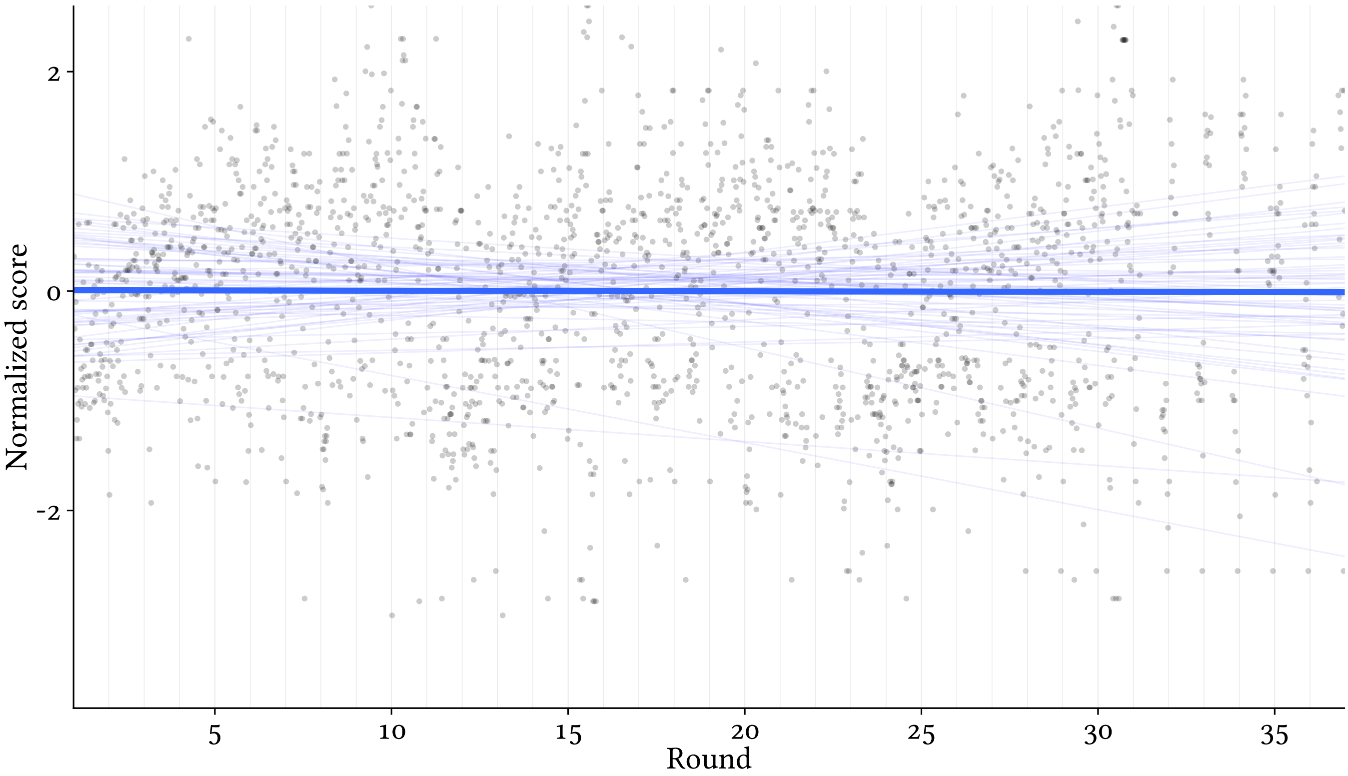


Figure 3.8.: Normalized scores over time. Each of the dots corresponds to a normalized score. Scores are normalized within sessions. Each of the thin lines correspond to the score development of a single participant. The thick line corresponds to the average development of scores. Scores are calculcated using the quadratic scoring rule 3.1.2.

Table 3.2.: Regression showing the impact of repeated use on scores.

	Scores$_{i,t}$				
	(I)	(II)	(III)	(IV)	(V)
t	-0.51	0.04	0.03	-1.53	-1.49
	(0.47)	(0.45)	(0.69)	(0.79)	(0.88)
t^2	0.02	-0.00	-0.00	0.00	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$p_{t-1} - f_{t-1}$			-0.11^{***}	-0.07^{**}	-0.06^{*}
			(0.02)	(0.02)	(0.03)
Δp_{t-1}			0.13^{**}	0.03	0.05
			(0.04)	(0.04)	(0.04)
$p_{t-1} - \bar{r}_{i,t-1}$			0.02	0.05	0.05
			(0.04)	(0.04)	(0.03)
Participant FE		✓	✓	✓	✓
Session FE				✓	✓
Market FE				✓	✓
Round FE					✓
Num. obs.	1944	1944	1656	1656	1656
R^2	0.00	0.19	0.23	0.28	0.28
Adj. R^2	0.00	0.17	0.20	0.25	0.25

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Std. errors clustered on participants.

the overall trend, which gives equal weight to each observation. The thin, gray lines are the trend lines for each of the 60 participants. The main message is that, on average, participants do not improve at forecasting[9]. One interpretation of this result is that the learning curve for the module is not steep. This is a desirable quality, as it implies that one can interpret the participants' expectations equally, irrespective of where in the time spectrum they were elicited.

To further investigate whether there is a learning effect over time we can incorporate characteristics of the experiment in a regression analysis. Regressions are shown in Table 3.2. The dependent variable in the regression table is individual scores as computed by the quadratic scoring rule.

The first column (I) in Table 3.2 displays a simple relationship between the round number and the scores. The result shows that is no relationship between the timing and the score. Column (II) adds participant fixed effects and the results are unchanged. Column (III) adds

$\hat{x}_j = (x_j - \bar{x}) \big/ \sqrt{(N-1)^{-1} \sum_i (x_i - \bar{x})^2}$.

9 As explained before, participants were required to complete two sets of understanding questions before starting the experiment; these questions acted as a tutorial.

additional explanatory variables. Firstly, the bias, which is the difference between the realized equilibrium price and the fundamental value, $p_{t-1} - f_{t-1}$, measures how far the price is from the rational equilibrium path. The second control is Δp_{t-1}, which is the change in the price in the last period compared with the period before that, thus capturing how fast the price is changing. Finally, $p_{t-1} - \bar{r}_{t-1}$ measures the difference between the realized price in the previous period and the mean of participant i's last expectation distribution. This is a measure of how far off the participant was in the previous period. While two of these measures are statistically significant, there are still no direct timing effects, as shown by the statistically insignificant result with respect to the timing variables. Column (IV) and Column (V) add additional fixed effects. Again, the result suggests that the number of times does not affect the precision.

The fact that the relationship between the round number and the score is negative may have to do with the underlying experiment, in which bubbles tend to occur towards the end of the markets. There is no evidence of any significant dynamic effects, as shown by the coefficient on t^2.

Overall, the regression suggests that there are few learning effects in repeated use of the `forecast.js` module.

3.4. Conclusion

This paper introduces a new module for eliciting expectations from participants in economic experiments. It enables researchers to elicit full subjective probability mass functions of individual expectations. Using real data, we show that participants tend to get faster at using the module with repeated usage. This suggests that experimenters may use it to elicit expectations several times, without adding too much time. We also show that participants' scores do not tend to change significantly over time, suggesting that they are not improving at making predictions.

The module can be easily incorporated in HTML-based frameworks, such as oTree. It can also be adapted to individual experimenters' needs.

Bibliography

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. 2017. *When Should You Adjust Standard Errors for Clustering.* NBER working paper 13565. Cambridge, MA: National Bureau of Economic Research, October.

Abreu, Dilip, and Markus K. Brunnermeier. 2003. "Bubbles and Crashes." *Econometrica* 71 (1): 173–204.

Aguirregabiria, Victor, and Chun-Yu Ho. 2012. "A dynamic oligopoly game of the US airline industry: Estimation and policy experiments." *Journal of Econometrics* 168 (1): 156–173. http://www.sciencedirect.com/science/article/pii/S030440761100176X.

Alpert, Marc, and Howard Raiffa. 1982. "A progress report on the training of probability assessors."

Aragón, Nicolás, and Rasmus Pank Roulund. 2019. *Certainty and Decision-Making in Experimental Asset Markets.* Thesis chapter. European University Institute. http://pank.eu/experiment.

Armantier, Olivier, and Nicolas Treich. 2013a. "Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging." *European Economic Review* 62:17–40.

———. 2013b. "Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging." *European Economic Review* 62:17–40.

Baghestanian, Sascha, Volodymyr Lugovskyy, and Daniela Puzzello. 2015. "Traders' heterogeneity and bubble-crash patterns in experimental asset markets." *Journal of Economic Behavior & Organization* 117:82–101.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2015. *Measuring Economic Policy Uncertainty.* Working paper. Economic Policy Uncertainty, March. http://www.policyuncertainty.com/media/BakerBloomDavis.pdf.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A model of investor sentiment." *Journal of financial economics* 49 (3): 307–343.

Benoît, Jean-Pierre, and Juan Dubra. 2011. "Apparent overconfidence." *Econometrica* 79 (5): 1591–1625.

Berry, Steven. 1994. "Estimating Discrete Choice Models of Product Differentiation." rand *Journal of Economics* 25 (2): 242–262.

Berry, Steven, Michael Carnall, and Pablo T. Spiller. 1996. *Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity.* Working Paper Series 5561. National Bureau of Economic Research, June. http://www.nber.org/papers/w5561.

Berry, Steven, and Panle Jia. 2010. "Tracing the Woes: An Empirical Analysis of the Airline Industry." *American Economic Journal: Microeconomics* 2, no. 3 (September): 1–43. http://www.aeaweb.org/articles.php?doi=10.1257/mic.2.3.1.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63, no. 4 (June): 841–890.

Bhattacharyya, A. 1943. "On a measure of divergence between two statistical populations defined by probability distributions." *Bulletin of the Calcutta Mathematical Society* 35:99–109.

Biais, Bruno, Denis Hilton, Karine Mazurier, and Sébastien Pouget. 2005. "Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market." *The Review of economic studies* 72 (2): 287–312.

Bilotkach, Volodymyr, Yuriy Gorodnichenko, and Oleksandr Talavera. 2012. "Sensitivity of prices to demand shocks: A natural experiment in the San Francisco Bay Area." *Transportation Research Part A: Policy and Practice* 46 (7): 1137–1151. http://www.sciencedirect.com/science/article/pii/S0965856412000468.

Blanco, Mariana, Dirk Engelmann, Alexander K. Koch, and Hans-Theo Normann. 2010. "Belief elicitation in experiments: is there a hedging problem?" *Experimental Economics* 13 (4): 412–438.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77, no. 3 (May): 623–685.

Bloom, Nick, Stephen Bond, and John van Reenen. 2007. "Uncertainty and Investment Dynamics." *Review of Economic Studies* 74:391–415.

Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis. 2008. "Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters." *The Economic Journal* 118 (July): 1107–1127.

Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer. 2011. "$\mathbb{D}^3$: Data-Driven Documents." *IEEE Transactions on Visualization and Computer Graphics* 17 (12): 2301–2309. http://vis.stanford.edu/papers/d3.

Boswijk, H. Peter, Cars H. Hommes, and Sebastiano Manzan. 2007. "Behavioral heterogeneity in stock prices." *Journal of Economic Dynamics and control* 31 (6): 1938–1970.

Brock, William A., and Cars H. Hommes. 1997. "A rational route to randomness." *Econometrica:* 1059–1095.

Bureau of Economic Analysis. 2015. *Regional Economic Accounts: GDP by Metropolitan Area.* Statistical table. Washington, D.C.: U.S. Department of Commerce. `http://www.bea.gov/regional/`.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press.

Carlé, Tim A, Yaron Lahav, Tibor Neugebauer, and Charles N. Noussair. 2019. "Heterogeneity of beliefs and trade in experimental asset markets." *Journal of Financial and Quantitative Analysis* 54 (1): 215–245.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9 (March): 88–97.

Chien, Tim, and Robert Nyman. 2013. *CSS Length Explained.* Mozilla Hacks. Mozilla. `https://hacks.mozilla.org/2013/09/css-length-explained/`.

Deaves, Richard, Erik Lüders, and Guo Ying Luo. 2008. "An experimental test of the impact of overconfidence and gender on trading activity." *Review of Finance* 13 (3): 555–575.

Diebold, Francis X., Anthony S. Tay, and Kenneth F. Wallis. 1997. *Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters.* Working Paper Series 6228. National Bureau of Economic Research, October.

Dixit, Avinash K., and Robert S. Pindyck. 1994. *Investment Under Uncertainty.* Princeton University Press.

Dubé, Jean-Pierre, Jeremy Fox, and Che-Lin Su. 2012. "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation." *Econometrica* 80, no. 5 (December): 2231–2267.

Engelberg, Joseph, Charles F. Manski, and Jared Williams. 2009. "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters." *Journal of Business & Economic Statistics* 27 (1): 30–41.

Escobari, Diego. 2012. "Dynamic Pricing, Advance Sales, and Aggregate Demand Learning in Airlines." *Journal of Industrial Economics* 60, no. 4 (December): 697–724.

———. 2014. "Estimating Dynamic Demand for Airlines." *Economics Letters* 124, no. 1 (July): 26–29.

Fisher, Kenneth L., and Meir Statman. 2000. "Cognitive biases in market forecasts." *The Journal of Portfolio Management* 27 (1): 72–81.

Hanaki, Nobuyuki, Eizo Akiyama, and Ryuichiro Ishikawa. 2018. "Effects of different ways of incentivizing price forecasts on market dynamics and individual decisions in asset market experiments." *Journal of Economic Dynamics and Control* 88:51–69.

Harrison, Glenn W., Jimmy Martínez-Correa, J. Todd Swarthout, and Eric Ulm. 2017. "Scoring Rules for Subjective Probability Distributions." *Journal of Economic Behavior & Organization* 134 (February): 430–448.

Harrison, Glenn W., Jimmy Martínez-Correa, J. Todd Swarthout, and Eric R. Ulm. 2015. "Eliciting subjective probability distributions with binary lotteries." *Economics Letters* 127:68–71.

Harrison, Glenn W., and Richard D. Phillips. 2014. "Subjective Beliefs and Statistical Forecasts of Financial Risks: The Chief Risk Officer Project." In *Contemporary Challenges in Risk Management: Dealing with Risk, Uncertainty and the Unknown,* edited by Torben Juul Andersen and Stanley Mayes, 163–202. Palgrave Macmillan.

Harrison, Glenn W., and Eric R. Ulm. 2015. *Recovering Subjective Probability Distributions.* Working Paper 01. Robinson College of Business, Georgia State University: Center for the Economic Analysis of Risk.

Harrison, J. Michael, and David M. Kreps. 1978. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations." *The Quarterly Journal of Economics* 92 (2): 323–336.

Haruvy, Ernan, Yaron Lahav, and Charles N. Noussair. 2007. "Traders' Expectations in Asset Markets: Experimental Evidence." *American Economic Review* 97 (5): 1902–1920.

Heemeijer, Peter, Cars Hommes, Joep Sonnemans, and Jan Tuinstra. 2009. "Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation." *Journal of Economic dynamics and control* 33 (5): 1052–1072.

Hommes, Cars. 2011. "The heterogeneous expectations hypothesis: Some evidence from the lab." *Journal of Economic dynamics and control* 35 (1): 1–24.

Hommes, Cars, et al. 2017. "Booms, busts and behavioural heterogeneity in stock prices." *Journal of Economic Dynamics and Control* 80:101–124.

Hommes, Cars, Joep Sonnemans, Jan Tuinstra, and Henk Van De Velden. 2008. "Expectations and bubbles in asset pricing experiments." *Journal of Economic Behavior & Organization* 67 (1): 116–133.

Hopenhayn, Hugo, and Richard Rogerson. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy* 101 (5): 915–938.

Hussam, Reshmaan N., David Porter, and Vernon L. Smith. 2008. "Thar She Blows: Can Bubbles Be Rekindled with Experienced Subjects?" *American Economic Review* 98, no. 3 (June): 924–37.

Ishii, Jun, Sunyoung Jun, and Kurt Van Dender. 2009. "Air travel choices in multi-airport markets." *Air travel choices in multi-airport markets* 65, no. 2 (March): 216–227.

Karni, Edi. 2014. “Axiomatic Foundations of Expected Utility and Subjective Probability.” Chap. 1 in *Handbook of the Economics of Risk and Uncertainty,* edited by Mark J. Machina and W. Kip Viscusi, vol. 1. North Holland.

Kellogg, Ryan. 2014. “The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling,” *American Economic Review* 104 (6): 1698–1734.

Kirchler, Erich, and Boris Maciejovsky. 2002. “Simultaneous over- and underconfidence: Evidence from experimental asset markets.” *Journal of Risk and Uncertainty* 25 (1): 65–85.

Kirchler, Michael, Jürgen Huber, and Thomas Stöckl. 2012. “Thar She Bursts: Reducing Confusion Reduces Bubbles.” *American Economic Review* 102, no. 2 (April): 865–83.

Knotek, Edward S. 2011. “Convenient Prices and Price Rigidity: Cross-Sectional Evidence.” *The Review of Economics and Statistics* (August): 1076–1086.

Lin, Lawrence I-Kuei. 1989. “A Concordance Correlation Coefficient to Evaluate Reproducibility.” *Biometrics* 45 (1): 255–268.

Manski, Charles F. 2004. “Measuring Expectations.” *Econometrica* 72, no. 5 (September): 1329–1376.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory.* Oxford University Press.

Matheson, James E., and Robert L. Winkler. 1976. “Scoring Rules for Continuous Probability Distributions.” *Management Science* 22 (10): 1087–1096.

McFadden, Daniel L. 1981. “Structural Discrete Probability Models Derived from Theories of Choice.” In *Structural Analysis of Discrete Data and Econometric Applications,* vol. Manski, Charles F. and McFadden, Daniel L. Cambridge, MA: The MIT Press. `http://elsa.berkeley.edu/~mcfadden/discrete.html`.

Michailova, Julija, and Ulrich Schmidt. 2016. “Overconfidence and bubbles in experimental asset markets.” *Journal of Behavioral Finance* 17 (3): 280–292.

Milgrom, Paul, and Nancy Stokey. 1982. “Information, trade and common knowledge.” *Journal of Economic Theory* 26 (1): 17–27.

Miller, Edward M. 1977. “Risk, uncertainty, and divergence of opinion.” *The Journal of Finance* 32 (4): 1151–1168.

Minsky, Hyman. 1992. *The Financial Instability Hypothesis.* Working Paper 74. Levy Economics Institute.

Moore, Don A., and Paul J. Healy. 2008. “The trouble with overconfidence.” *Psychological review* 115 (2): 502.

Nevo, Aviv. 2001. “Measuring Market Power in the Ready-to-Eat Cereal Industry.” *Econometrica* 69, no. 2 (March): 307–342.

Nevo, Aviv, and Michael D. Whinston. 2010. "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives* 24 (2): 69–82.

Noussair, Charles N., and Steven Tucker. 2013. "Experimental Research on Asset Pricing." *Journal of Economic Surveys* 27:554–569.

Palan, Stefan. 2013. "A Review of Bubbles and Crashes in Experimental Asset Markets." *Journal of Economic Surveys* 27 (3): 570–588.

Plott, Charles R., and Shyam Sunder. 1988. "Rational expectations and the aggregation of diverse information in laboratory security markets." *Econometrica: Journal of the Econometric Society:* 1085–1118.

Powell, Owen, and Natalia Shestakova. 2016. "Experimental asset markets: A survey of recent developments." *Journal of Behavioral and Experimental Finance* 12:14–22.

Research and Innovative Technology Administration. 2013. *Airline Origin and Destination Survey* db1b. Washington, DC: U.S. Department of Transportation.

Reynaerts, Jo, Ravi Varadhan, and John C. Nash. 2012. *Enhancing the Convergence Properties of the* BLP *(1995) Contraction Mapping.* Vives discussion paper series 35. Katholieke Universiteit Leuven, Faculteit Economie en Bedrijfswetenschappen, Vives, August.

Roland, Christophe, Constantine Frangakis, and Ravi Varadhan. 2007. "Squared polynomial extrapolation methods with cycling: an application to the positron emission tomography problem." *Numerical Algorithms* 44, no. 2 (February): 159–172.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55, no. 5 (September): 999–1033.

Rutström, E. Elisabet, and Nathaniel T. Wilcox. 2009. "Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test." *Games and Economic Behavior* 67 (2): 616–632.

Savage, Leonard J. 1954. *The Foundations of Statistics.* New York: John Wiley.

———. 1971. "Elicitation of Personal Probabilities and Expectations." *Journal of the American Statistical Association* 66 (336): 783–801.

Scheinkman, Jose A., and Wei Xiong. 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy* 111 (6): 1183–1220.

Skrainka, Benjamin S. 2012. "Three essays on product differentiation: computational tools for applied research, evaluating model behavior, and geographic demand." PhD diss., University College London. http://discovery.ucl.ac.uk/1344087/1/1344087.pdf.

Smith, Vernon L., Gerry L. Suchanek, and Arlington W. Williams. 1988. "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica* 56 (5): 1119–1151.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economic Statistics* 20, no. 4 (October): 518–529.

Stöckl, Thomas, Jürgen Huber, and Michael Kirchler. 2010. "Bubble measures in experimental asset markets." *Experimental Economics* 13 (3): 284–298.

Sunder, Shyam. 1995. "Experimental Asset Markets: A Survey." Chap. 6 in *Handbook of Experimental Economics,* edited by Alvin E. Roth and John Kagel, 445–500. New Jersey: Princeton University Press.

Tirole, Jean. 1982a. "On the possibility of speculation under rational expectations." *Econometrica: Journal of the Econometric Society:* 1163–1181.

———. 1982b. "On the Possibility of Speculation under Rational Expectations." *Econometrica* 50 (5): 1163–1181.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation.* 2nd ed. Cambridge University Press.

Varadhan, Ravi. 2012. squarem*: Squared extrapolation methods for accelerating fixed-point iterations.* R package version 2012.7-1. `http://CRAN.R-project.org/package=SQUAREM`.

Varian, Hal R. 1985. "Divergence of opinion in complete markets: A note." *The Journal of Finance* 40 (1): 309–317.

Zarnowitz, Victor, and Louis A. Lambros. 1987. "Consensus and Uncertainty in Economic Prediction." *Journal of Political Economy* 95, no. 3 (June): 591–621.

Appendix A.

Appendix to Chapter 1

A.1. Further robustness checks

A.1.1. Additional graph for Hypothesis 2

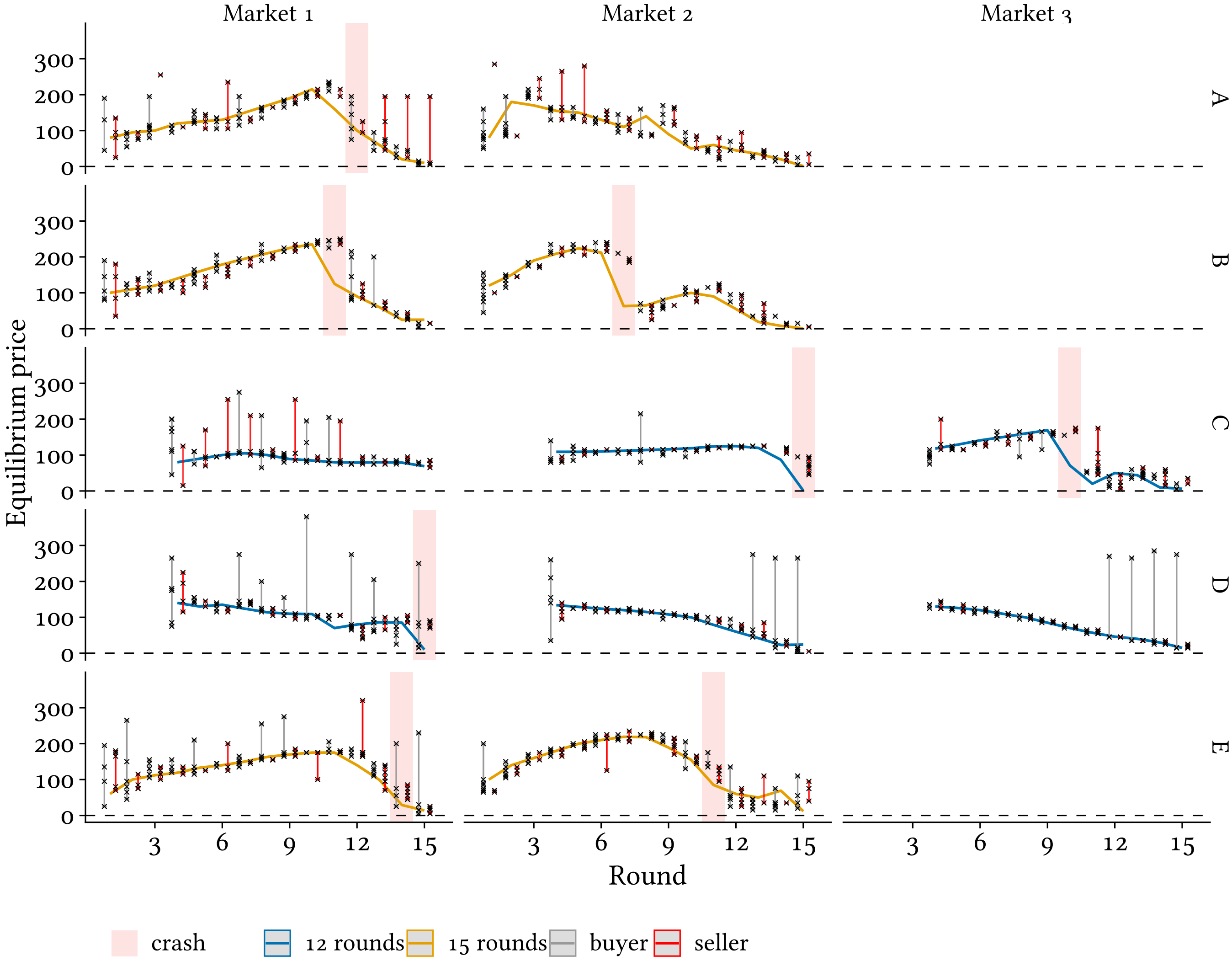


Figure A.1.: Price expectations of individual buyers and sellers. Each cross marks the mean expectation of a buyer or seller.

A.1.2. Increased agreement with the Bhattacharyya coefficient

As a further robustness check of Hypothesis 2 we measure the the degree of divergence of traders' belief distributions using the Bhattacharyya coefficient (Bhattacharyya, 1943). For discrete random variables it is defined as

$$BC(q_1, q_2) = \sum_{i=1}^{N} \sqrt{q_{1,i} q_{2,i}}, \tag{A.1}$$

where q_1 and q_2 are price forecast distributions for two different traders. Here, $q_{1,j} \in q_1$ is the probability that the realized price falls in price bracket j in accordance with the forecast of the first trader. Note that if the two probability distributions assign equal mass to all brackets then $BC(p, q) = \sum_{i=1}^{K} \sqrt{q_{1,i} q_{2,i}} = \sum_{i=1}^{n} \sqrt{q_{1,i}^2} = 1$. Likewise, if the two distributions are completely disjoint such that there exists no j such that $q_{1,j} > 0$ and $q_{2,j} > 0$ then $BC(p, q) = 0$. As such, $0 \leq BC(q_1, q_2 \leq 1$ and a higher BC value suggests less divergence between the two distributions, q_1 and q_2. Unlike Lin's concordance index, the Bhattacharyya coefficient does take into account the moments of a distribution.

To test whether beliefs converge over time we calculate the BC coefficient for all pairs of traders in each round of the experiment. Thus, we have 132 BC coefficients per round.

As a first test to address Hypothesis 2, we check whether the observed overlap of forecasts is similar. We perform the Wilcoxon Signed Rank Sum Test with a one-sided alternative hypothesis. Specifically, for market $m = 1, 2$ we compare the difference $BC_{ij}^{m} - BC_{ij}^{m+1}$ for traders i and j and test that there is no difference in the median under the null. We use a non-parametric test over a t-test, as the distributions of differences between values of market 1 and market 2 have fat tails and are spread over the domain $[-1, 1]$. Note that for the comparison between market 2 and 3 we are only using data from the sessions with 12 rounds per market.

In the last row of Table A.1, denoted *All*, we show the median overlap in market 1, 2 and 3 using data from all sessions. The numbers in the parentheses are the standard deviation and the stars denote the p-value of the Wilcoxon Signed Rank Sum Test. As can be seen, the median level of overlap in forecasts is increasing over the markets. The most relevant comparison is between market 1 and market 2, as they both use the full sample of data. As the table show, the median overlap over all sessions increases by approximately 0.05 between the two markets. The exact p-value of the test is less than 0.001, and, as such, we strongly reject the null that the median is the same.

The upper part of Table A.1 compares median overlaps of predictions across rounds and sessions. The main difference across markets is the starting level of overlap of beliefs, which is increasing across markets. In the sample, the median overlap of expectations in the first round is 0.158 in the first market, 0.450 in the second market and 0.667 in the third market.

Table A.1.: Average overlap of distributions as measured by BC. Plain numbers denote the median while numbers in parentheses denote the standard deviation. The superscripted stars denote the significance of the p-value of the Wilcoxon Signed Rank Sum Test between the observations from period j in markets i and $i-1$.

t	1		2		3	
1	0.260	(0.285)	0.453^{***}	(0.319)		
2	0.554	(0.303)	0.655^{***}	(0.241)		
3	0.652	(0.253)	0.675^{**}	(0.303)		
4	0.522	(0.337)	0.589^{**}	(0.315)	0.593	(0.311)
5	0.566	(0.318)	0.680^{***}	(0.282)	0.785^{***}	(0.261)
6	0.580	(0.336)	0.690^{***}	(0.277)	0.782^{**}	(0.267)
7	0.614	(0.345)	0.705^{***}	(0.293)	0.695	(0.310)
8	0.634	(0.308)	0.639	(0.324)	0.683	(0.362)
9	0.692	(0.271)	0.639	(0.289)	0.707	(0.312)
10	0.641	(0.349)	0.636	(0.303)	0.689	(0.283)
11	0.660	(0.321)	0.625	(0.327)	0.570	(0.365)
12	0.511	(0.350)	0.608^{***}	(0.315)	0.689	(0.322)
13	0.551	(0.329)	0.704^{***}	(0.283)	0.764^{**}	(0.286)
14	0.577	(0.330)	0.669^{***}	(0.295)	0.660	(0.338)
15	0.626	(0.339)	0.584	(0.334)	0.649	(0.317)
All	0.576	(0.334)	0.637^{***}	(0.306)	0.689	(0.318)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A.2.: Regression analysis on overlap of expectations.

	$BC_{c,ij} - \bar{BC}_{c,ij}$				
	(I)	(II)	(III)	(IV)	(V)
Market 2	0.057^{*}	0.043	0.044	0.044	0.044
	(0.026)	(0.026)	(0.026)	(0.025)	(0.025)
Market 3	0.105^{*}	0.080^{*}	0.050	0.052	0.052
	(0.042)	(0.035)	(0.049)	(0.047)	(0.047)
$p_{t-1} - f_{t-1}$		−0.000	−0.000	0.000	0.000
		(0.000)	(0.000)	(0.001)	(0.001)
Session FE			✓	✓	✓
Round FE				✓	✓
Participant FE					✓
Num. obs.	10692	9900	9900	9900	9900
R^2	0.014	0.011	0.025	0.034	0.289
Adj. R^2	0.014	0.010	0.024	0.032	0.284

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Std. errors clustered on IDs and sessions.

Table A.3.: Regression results of (1.3) using $\sigma_{m,t}$ instead of $\sigma^{m}_{m,t}$.

	Δp_t				
	(I)	(II)	(III)	(IV)	(V)
Δp_{t-1}	0.33^{***}	-0.24	-0.26	-0.17	-0.22
	(0.07)	(0.22)	(0.20)	(0.20)	(0.23)
$p_{t-1} - f_{t-1}$	-0.29^{***}	-0.25^{***}	-0.29^{***}	-0.26^{***}	-0.28^{***}
	(0.05)	(0.05)	(0.04)	(0.05)	(0.06)
σ_t	-0.96^{*}	-0.65	-0.88^{*}	-0.96^{*}	-1.85^{**}
	(0.44)	(0.44)	(0.42)	(0.41)	(0.58)
ΔE_t		0.60^{**}	0.29	0.69^{**}	0.89^{**}
		(0.21)	(0.21)	(0.25)	(0.27)
N_t			-31.57^{***}	-26.35^{***}	-17.25
			(7.32)	(7.34)	(9.09)
$E_{t-1} - p_{t-1}$				-0.58^{**}	-0.69^{**}
				(0.20)	(0.23)
Session FE					✓
Market FE					✓
Round FE					✓
Num. obs.	138	138	138	138	138
R^2	0.31	0.35	0.43	0.47	0.58
Adj. R^2	0.30	0.33	0.41	0.44	0.49

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ Robust std. errors.

Table A.2 shows a regression analysis of the data. The dependent variable is BC_{ij} and we control the regressions with a number of fixed effects. As the table shows, the overlap measure increases by 0.06 from market 1 to 2, and approximately the same from market 2 to 3, although the significance of the latter is decreasing. Note that in column (IV) we control for participant, session and round.

A.1.3. Additional robustness checks for Hypothesis 3

Table A.4.: Regression results of (1.3) using $\sigma_{m,t}$ instead of $\sigma^m_{m,t}$, using only observations from before a crash.

	Δp_t				
	(I)	(II)	(III)	(IV)	(V)
Δp_{t-1}	0.43^{***}	0.19	−0.01	0.05	−0.17
	(0.08)	(0.20)	(0.19)	(0.18)	(0.22)
$p_{t-1} - f_{t-1}$	-0.12^{***}	-0.11^{***}	-0.15^{***}	-0.12^{***}	−0.09
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)
σ_t	-0.96^{**}	-0.85^{*}	-0.78^{*}	-0.84^{**}	-1.43^{**}
	(0.33)	(0.34)	(0.31)	(0.30)	(0.48)
ΔE_t		0.24	0.03	0.32	0.45^{*}
		(0.18)	(0.17)	(0.20)	(0.22)
N_t			-26.45^{***}	-25.19^{***}	-17.68^{*}
			(5.70)	(5.55)	(6.86)
$E_{t-1} - p_{t-1}$				-0.47^{**}	-0.44^{*}
				(0.17)	(0.20)
Session FE					✓
Market FE					✓
Round FE					✓
Num. obs.	105	105	105	105	105
R^2	0.28	0.29	0.42	0.46	0.59
Adj. R^2	0.26	0.27	0.39	0.43	0.46

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ Robust std. errors.

A.2. Instructions for experiment [1]

A.2.1. General Instructions

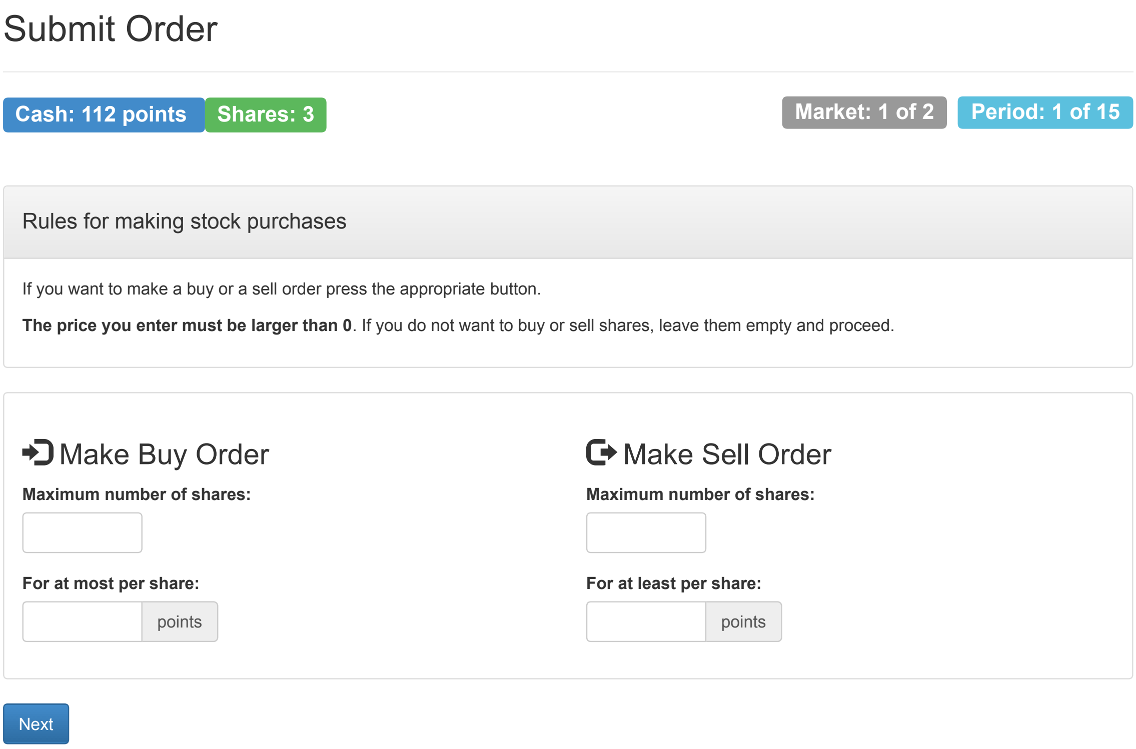
This is an experiment in the economics of market decision making. The instructions are simple and if you follow them carefully and make good decisions, you might earn a considerable amount of money, which will be paid to you in cash at the end of the experiment. The experiment will consist of two sequences of 15 trading periods in which you will have the opportunity to buy and sell in a market. The currency used in the market is “points”. All trading will be in terms of points. The cash payment to you at the end of the experiment will be in euros. The conversion rate is 85 points to 1 dollar.

1 These instructions are those from Haruvy et al. (2007), with an added section explaining the novel forecast tool, and a different mechanism to deliver the payment.

A.2.2. How to use the computerized market

General Instructions for asset trading

In each period, you will see a computer screen like the one shown below.



You can use the interface to buy and sell Shares. At the top of your computer screen, in top left corner, you can see the Money and Shares you have available.

At the beginning of each trading period, if you wish to purchase shares you can send in a buy order. Your **buy order** indicates the number of shares you would like to buy and the highest price that you are willing to pay. Similarly, if you wish to sell shares, you can send in a **sell order**. Your sell order indicates the number of shares you are offering to sell and the lowest price that you are willing to accept. The price at which you offer to buy must be less than the price at which you offer to sell. The price you specify in your order is a per-unit price, at which you are offering to buy or sell each share. In each period you can, if you wish, do simultaneously a buy and a sell order at different prices.

The computer program will organize the buy and sell orders and uses them to determine the **trading price** at which units are bought and sold. All transactions in a given period will occur at the same trading price. This will generally be a price where the number of shares with sell order prices at or below this clearing price is equal to the number of shares with buy order prices at or above this clearing price. The people who submit buy orders at prices above the trading price make purchases, and those who submit sell orders at prices below the trading price make sales.

Example of how the market works: Suppose there are four traders in the market and:

- Trader 1 submits an offer to buy at 60
- Trader 2 submits an offer to buy at 20
- Trader 3 submits an offer to sell at 10
- Trader 4 submits an offer to sell at 40

At any price above 40, there are more units offered for sale than for purchase. At any price below 20 there are more units offered for purchase than for sale. At any price between 21 and 39 there is an equal number of units offered for purchase and for sale. The trading price is the lowest price at which there is an equal number of units offered for purchase and for sale. In this example that price is 21. Trader 1 makes a purchase from trader 3 at a price of 21.

Specific Instructions for this Experiment

The experiment will consist of two independent sequences of 15 trading periods. In each period, there will be a market open, operating under the rules described above, in which you are permitted to buy and sell shares.

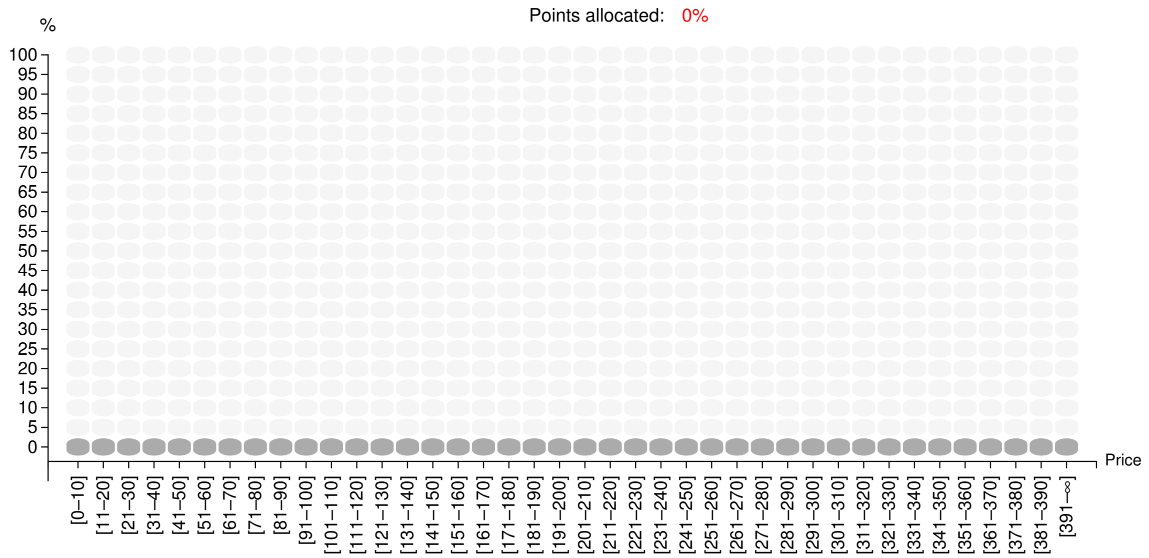
Shares have a life of 15 periods. Your shares carry over from one trading period to the next. For example, if you have 5 shares at the end of period 1, you will have 5 shares at the beginning of period 2.

You receive dividends for each share in your inventory at the end of each of the 15 trading periods. At the end of each trading period, including period 15, each share you hold will pay you a dividend of 0, 4, 14, or 30, each with equal chance. This means that the average dividend for each share in each period is 12. The dividend is added to your money balance automatically after each period. After the dividend is paid at the end of period 15, the market ends and there are no further earnings possible from shares in the current market.

A new 15-period market will then begin, in which you can trade shares of a new asset for 15 periods. The amount of shares and money that you have at the beginning of the new market will be the same as at the beginning of the first 15 period market. There will be two 15 period markets making up the experiment.

Making Predictions

In addition to the money you earn from dividends and trading, you can make money by accurately forecasting the trading prices of all future periods. You will indicate your forecasts using the prediction tool shown below.

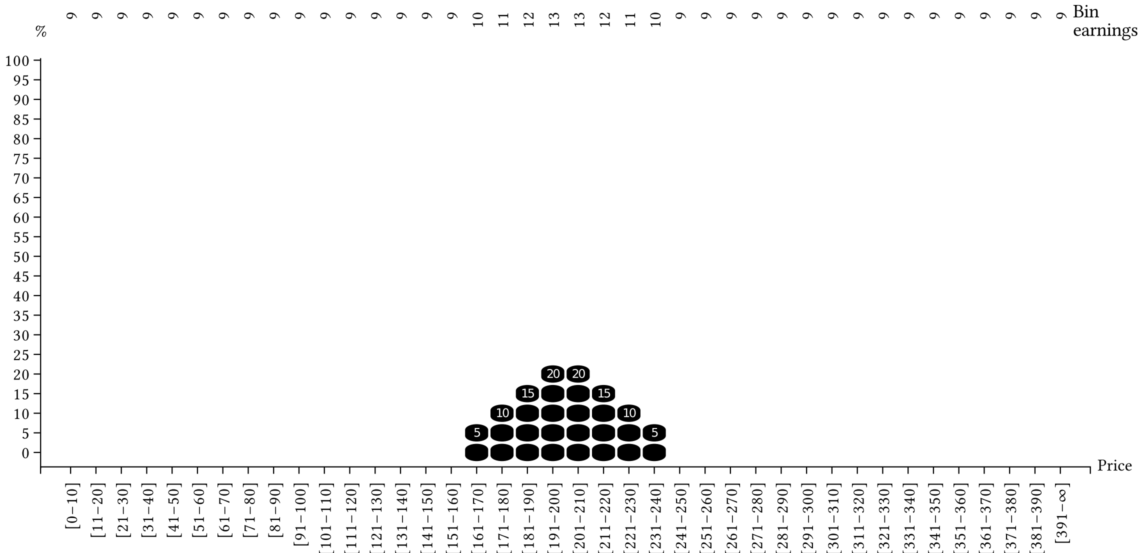


Predictions are produced in the following way. The range of prices is divided in intervals of 10, except for the last one that is valid for all prices greater than 390. You have 20 tokens, that you can put in the different prices that you believe will be next period's realized price.

In the bottom part of the prediction tool, you can find several pre-set distributions which you can use to speed up the process. You can, also, drag the distributions with the mouse or add and remove tokens by clicking. You will not be allowed to move forward until you assign all your tokens. Once you have assigned all of them, you will click in "Finalize" and then in "Next".

In the upper part of the screen you can see market prices for the previous periods.

The money you receive from your forecasts will depend on how many tokens you allocate on the effective price and how dispersed are your tokens. The more tokens you put in a price, the greater the payment. The more dispersed your tokens are, the lower the payment. The number of points that you will receive if the market price falls in a given interval is shown on the upper part of the tool after clicking “Finalize”.



Your Payment

Within each market you receive points by predictions and by transactions. The earnings by transactions is the number of points that you accumulated at the end of period 15, after the last dividend is paid. The earnings for predictions are the sum of the points that you obtained in each period of the market.

At the end of each 15-period market, it will be randomly chosen if you are paid by your earnings in predictions or in transactions, so you need to put effort in both tasks.

Therefore, your payment in a 15 period market will be one of the following options,

1. If you are chosen to be paid by transactions,

   The money you have at the beginning of period 1
   + the dividends you receive
   + the money received from sales of shares
   – the money spent on purchases of shares
   Converted to euros at 85 points per euro,

or

2. If you are chosen to be paid for your forecasts,

   The sum of the earnings from all forecasts for the 15 periods in a market, transformed into cash at the end of each market at 85 points per euro.

A.3. Questionnaire

A.3.1. Before Session

Understanding questions

Prediction Tool

Please, make predictions using the tool. You have to allocate the tokens to the values that you think will be the values tomorrow. For example, if you think that the price tomorrow is going to be 5, 10 or 50, you have to put tokens on those values. Do not forget to click "Finalize" after you are done!

1. Put 25% of tokens on the price 96, 10% on price 106, 50% on price 198, and 15% on price 210.

This means that you think the price 198 is the most likely result that happens with a 50% probability, the second price you think is more likely to happen is 96 which happens with 25% probability and so on.

1. Now use the mouse to drag the distributions towards the sides, so that there is no tokens on prices 96, 106, 198 and 210.

Asset Trading

You have been assigned some shares. In this context, a share is a piece of paper that pays a return in each of the remaining periods, with a life of 15 periods. The share pays returns of {0, 4, 14, 30} with equal probability. Therefore, the return can be 14 in period 1, 0 in period 2, etc. The shares are worth 0 after finishing the last period.

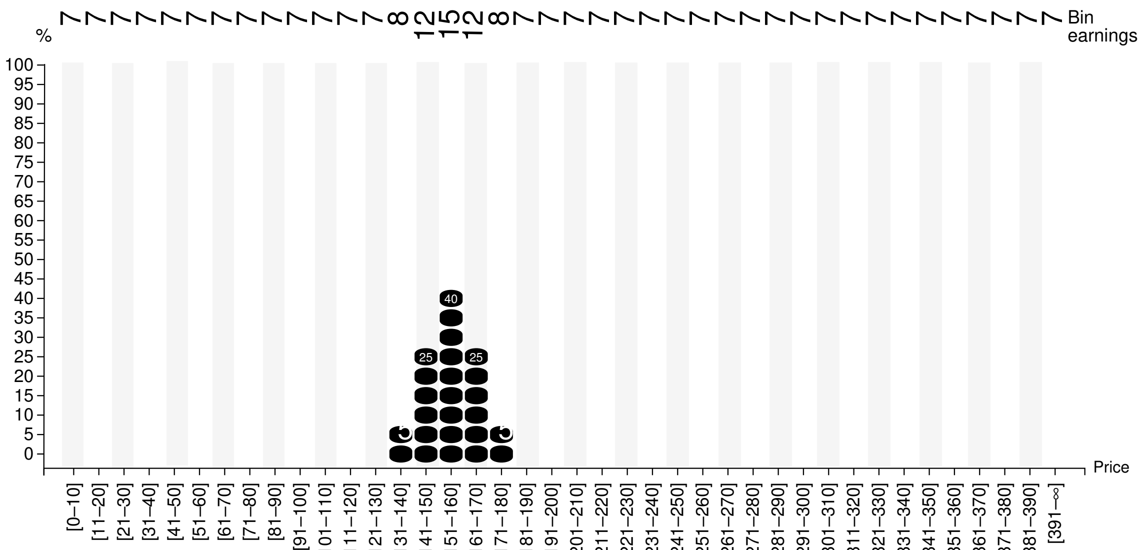
1. What is the average dividend payment on the share in period 3?
   Answer: 12 francs

2. What is the average total dividend that you will receive if you hold the share from period 3 through to the end of the market round (i.e. the end of period 15)?
Answer: 156 francs

3. What is the maximum possible dividend that you can receive if you hold the share from period 3 through to the end of the market round?
Answer: 390 francs

4. What is the minimum possible dividend that you can receive if you hold the share from period 3 through to the end of the market round?
Answer: 0 francs

Prediction Tool II



1. How many francs would you win if the price is 155?
Answer: 15

2. How many francs would you win if the price is 180?

Answer: 8

Cognitive Reflection Test

Payment randomly for a selected question at the end

(1) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does

the ball cost? _____ cents

Answer: 0.5

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes

Answer: 5

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ day

Answer: 47

A.3.2. After Session

Risk Aversion Lotteries

Which bet would you take?

a) 1/10 of $2.00 9/10 of $1.60 or 1/10 of $3.85, 9/10 of $0.10

b) 2/10 of $2.00, 8/10 of $1.60 or 2/10 of $3.85, 8/10 of $0.10

c) 3/10 of $2.00, 7/10 of $1.60 or 3/10 of $3.85, 7/10 of $0.10

d) 4/10 of $2.00, 6/10 of $1.60 or 4/10 of $3.85, 6/10 of $0.10

e) 5/10 of $2.00, 5/10 of $1.60 or 5/10 of $3.85, 5/10 of $0.10

f) 6/10 of $2.00, 4/10 of $1.60 or 6/10 of $3.85, 4/10 of $0.10

g) 7/10 of $2.00, 3/10 of $1.60 or 7/10 of $3.85, 3/10 of $0.10

h) 8/10 of $2.00, 2/10 of $1.60 or 8/10 of $3.85, 2/10 of $0.10

i) 9/10 of $2.00, 1/10 of $1.60 or 9/10 of $3.85, 1/10 of $0.10

j) 10/10 of $2.00, 0/10 of $1.60 or 10/10 of $3.85, 0/10 of $0.10

Exit Questionnaire

1. What is your gender?

[Female, Male]

2. What is your age in years?

[Age]

3. Your nationality?

[Open text]

4. What is your employment status?

[Full-time, Part-time, None]

5. If you are a student, what is your major?

[Open Text]

6. What is the level of the highest degree you are currently studying?

[Bachelor, Master, Doctor/PhD, Other]

17. Did you ever make a mistake in entering a price, or clicked a wrong button? If so, please tell us exactly what went wrong and in what period:

[Open text]

18. Did you find the instructions in the market experiment clear and understandable? What if anything could be improved?

[Open Text]

Appendix B.

Appendix to Chapter 3

B.1. Robustness check of precision

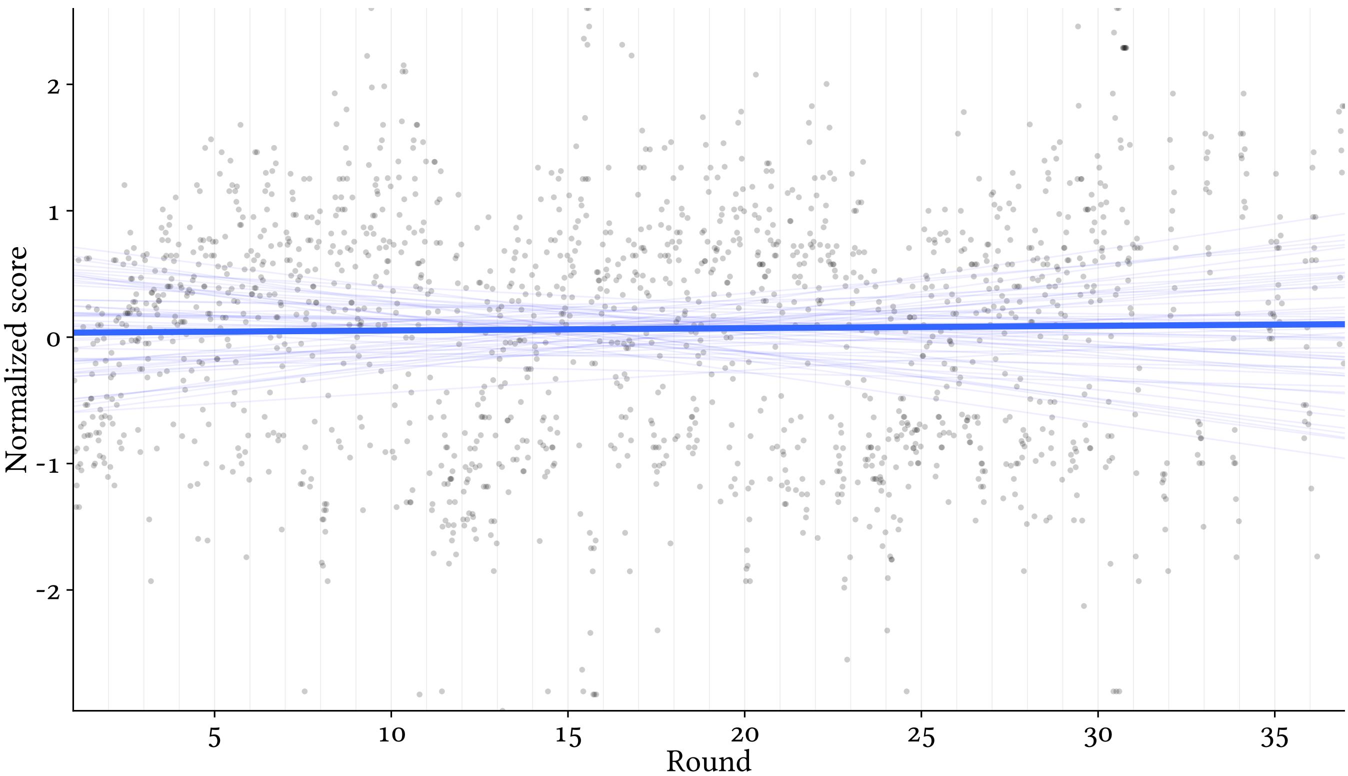


Figure B.1.: Normalized scores over time. Here, the seven participants with the lowest earning and the five participants with the highest earnings are excluded.

Figure B.1 replicates Figure 3.8 using only a subset of the data. We exclude the seven participants with the lowest overall earnings (corresponding to the lowest 10%) and the five participants with the highest earnings (corresponding to the top 5%). Both groups are characterized by earning significantly more (less) than the remaining participants. As can be seen from the figure, excluding this data does not alter the overall conclusions posed above. The results are also similar if we only leave out the seven participants who performed the worst in the forecasting exercise.

B.2. Using `forecast.js` in a standalone HTML page

In this section we show a minimal example of how to include the forecast module on a plain HTML page.

```
% <!DOCTYPE html>
<html lang="en">
  <head>

    <meta charset="utf-8">
    <title>Forecast.js elictitation module</title>

    <!-- Load the CSS styles needed for forecast.js -->
    <link rel="stylesheet" href="forecast-js.css"/>

    <!-- forecast.js depends on d3.js.   -->
    <!--I assume both are available in the "js" folder -->
    <script type="text/javascript" src="/js/d3.v3.min.js"></script>
    <script type="text/javascript" src="/js/forecast.js"></script>

  </head>

  <body>

    <!-- First, a block with usage instructions -->

    <div id="instructions">
      <h2>Price prediction instructions</h2>
      <p>
        Use the forecast tool to predict future prices.
      </p>
      <p>
        Make distrbutions by clicking on the gray boxes or by using
        one of the predefined distribution available via the buttons
        under the tool.
      </p>
      <p>
```

```
        After you have made a distribution you can
        change the location by dragging it with the mouse.
      </p>
      <p>
        When you are satisfied with your choice click <em>Finalize</em>.
      </p>
      <p>
        Note that once you have clicked finalized, the score is
        available as a JSON string under the element with the
        IDs <code>dist</code> and <code>value</code>.
      </p>
    </div>

    <!-- The following will initialize the an instance of forecast.js -->
    <script>

      <!-- We can change the settings using pre-defined variables -->
      var xname = "Price";
      var xmax = 100, xbins = 10;
      var alpha = 1, beta = 1;

      <!--   Give the instance the id "forecast";  -->
      <!--   Saving the instance as element "forecast_instance"  -->
      forecast_instance = add_forecast_widget("forecast");
    </script>
  </body>
</html>
```

B.3. Using `forecast.js` with oTree

In this section we show a minimal example of how to include the forecast module in the experimental software suit oTree (Chen et al., 2016).

To initialize oTree, set up a new virtual oTree environment, make a new `oTree` folder, and install oTree. While the oTree manual explains this in detail[1], the following should suffice on UNIX-like systems:

```
% # First, create a new oTree environment and activate it
mkdir oTree; cd oTree
virtualenv3 env
```

1 Please refer to `https://otree.readthedocs.io/en/latest/install.html`

```
source env/bin/active
# Install the latest version of oTree
pip3 install -U otree-core
# Create a new oTree instance
otree startproject elicitation-project
cd elicitation-project
# Create a new "app".
otree startapp simple_elicitation
```

This initializes a new oTree installation in its own virtual environment and creates a new oTree "app" called `simple_elicitation`. This app contains three important elements: the files `models.py` and `pages.py`, as well as the `templates` folder.

First, create a new folder called `static` and copy `d3.v3.min.js`, `forecast.js` and `forecast.css` into this folder. The `static` folder holds images, style files and Javascript programs.[2]

B.3.1. Setting up `models.py`

Next, we look at the `models.py` file. The `models.py` file defines and structures the space in which the data is stored. For instance, it defines variables/columns that are to be created in the database that stores the data. To store the elicited forecasts from participants, we must thus create a field to store the elicitation in. The `forecast.js` module automatically stores the elicitation in a JSON array, stored as a string, as mentioned above. In the example below, the fields `pred` and `score` in the `player` class are examples of how the elicitation results can be stored.

The rest of the file is standard. The first part is imports and are mostly setup automatically by oTree itself. In addition, some useful libraries and functions are imported, starting with `random`.

In the Constants class of the file the variables for the `forecast.js` module can be set up. The Player class sets up the fields in which participants' data are stored. In particular, in this example, the fields `forecast.js` `pred` and `score` are stored per participant in the player class.

In the group class, we find the score and the correct number of points in the bin in which the realized price fell.

2 The Django documentation of the static folder can be found here:
`https://docs.djangoproject.com/en/1.11/howto/static-files`

```
% # -*- coding: utf-8 -*-
# <standard imports>
from __future__ import division
from otree.db import models
from otree.constants import BaseConstants
from otree.models import BaseSubsession, BaseGroup, BasePlayer
from otree import widgets
from otree.common import Currency as c, currency_range
from django.conf import settings
# </standard imports>

import random
import json
from bisect import bisect_left
from quantile import quantile
from django.core.validators import RegexValidator

author = 'Rasmus Pank Roulund and Nicolas Aragon'

doc = """
Showcase of forecast.js
"""
keywords = ("Forecasting", "Finance", "Elicitation", "Trade")


class Constants(BaseConstants):
    players_per_group = None

    ## Prediction parameters
    prediction_kappa = 5
    prediction_alpha = 6
    prediction_beta = 20

    xbins_n = 40
    xmax = 400
    xbins = range(0, xmax, xmax//xbins_n)
    ybins_n = 100//5

class Player(BasePlayer):
    # <built-in>
    subsession = models.ForeignKey(Subsession)
    group = models.ForeignKey(Group, null = True)
    # </built-in>
```

```python
    ## Elicitation score and number of points in right bin
    elicitaion_earning = models.CurrencyField(initial=c(0))
    points_in_realized_price = models.PositiveIntegerField(initial=0)

    pred = models.TextField(doc = "Allocation of prediction points",
                            validators=[RegexValidator(
                                regex='\[([0-9]+ ?,? ?)+\]')])
    score = models.TextField(doc = "Scores assuming price in bin",
                             validators=[RegexValidator(
                                 regex='\[([0-9]+ ?,? ?)+\]')])


class Group(BaseGroup):
    # <built-in>
    subsession = models.ForeignKey(Subsession)
    # </built-in>

    realized_price = max([random.normalvariate(150,50), 0])

    def update_price(self):
        """Update the price, allowing for structural breaks"""
        raise NotImplementedError

    def calculate_elicitation_earnings(self):
        """Update the earnings from predictions in previous periods.

        The predicted earnings are calculated for each round within
        the current market and is calculated based on the player's
        predictions.

        """

        pred_field = "pred"
        score_field = "score"

        # Find the index of the bin with the realized price:
        price = self.realized_price
        pbins = Constants.pbins
        N = Constants.ybins_n
        realized_price_index = max(0, bisect_left(pbins, price)-1)

        ## Now iterate over each player in the group and update the
        ## variables for
        ## - How many points they allocated correctly
```

```
        ## -
        for p in self.get_players():
            preds = json.loads(getattr(p, pred_field))
            scores = json.loads(getattr(p, score_field))

            p.point_in_realized_price = preds[realized_price_index]
            p.elicitaion_earning = scores[realized_price_index]
```

B.3.2. The pages.py file

To insert the module into the actual experiment, the module must be added to the pages.py file. An example is included below.

```
% # -*- coding: utf-8 -*-
from __future__ import division
from .\ import models
from ._builtin import Page, WaitPage
from .models import Constants
from django.utils.translation import ugettext as _

import json

##* PREDICTION PAGES

class Prediction(Page):
    """Pages for making predictions.

    Participants will make predictions for all future periods.

    """
    form_model = models.Player

    def get_form_fields(self):
        """Return a list of prediction fields to modify

        Actually, it is not necessary to dynamically make this list...
        """
        fields = ["pred"
                  # ,"custom_A", "custom_B"
                 ]
        return(fields)
```

```python
    def vars_for_template(self):
        variables = {
            "alpha"   : Constants.prediction_alpha
            "beta"    : Constants.prediction_beta
            "kappa"   : Constants.prediction_kappa
            "xmax"    : Constants.xmax
            "xbins_n" : Constants.xbins_n
            "ybins_n" : Constants.xbins_n}
        return(variables)

    def error_message(self, vals):
        """Make sure that predictions do not exceed 100%.

        Add check in case somebody is altering variables with the
        JavaScript console.

        """
        preds = {key: vals[key] for key in vals.keys()
                  if key.startswith("pred")}
        for pred in preds.values():
            p = sum(json.loads(pred))
            if p > Constants.ybins_n:
                return (_("You have assigned more than 100%!"))
            elif (p < Constants.ybins_n):
                return (_("Not all points have been assigned."))



class Results(Page):
    """Page shows the earnings in a single round of market."""

    # timeout_seconds = 45

    def vars_for_template(self):
        x = self.subsession.round_number
        m, r = Constants.interpret_round[x]
        change =  (self.group.equilibrium_price
                   * self.player.shares_change)
        temp_vars = {
            "total_cash_earning": (self.player.dividend_earnings
                                   + change),
            "asset_trading": change}
        if r >= 3:
            temp_vars["timelimit"] = 30
```

```
        return (temp_vars)

class OneMoreRoundP(Page):
    """Ask if the player wants one more round"""

    def is_displayed(self):
        return(False)

class FinalResult(Page):
    """Show the results of all """

    def is_displayed(self):
        return(False)

page_sequence = [Prediction,
                 Results,
                 OneMoreRoundP,
                 FinalResult]
```

B.3.3. Display forecast modules on the pages

The HTML template may be added using a code similar to the following example. This adds the forecast element to a page while using the experiment's favorite paramters.

If the page is saved as, for instance, `InsertForecast.html`, it can be "called" from other pages by issuing the following Django template command:

```
{% include 'asset_simple/InsertForecast.html' %}~

    {% load staticfiles otree_tags %}
    {% load staticfiles %}
    {% load i18n %}

    <!-- Hidden fields to store predictions -->
    {% for field in form %}
      {{ field.as_hidden }}
    {% endfor %}

    <script type="text/javascript"
            src="{% static 'asset_simple/d3.v3.min.js' %}">
    </script>
```

```
<link rel="stylesheet"
      href="{% static 'asset_simple/forecast_style.css' %}">

<script>
  var alpha = {{ alpha }};
  var beta = {{ beta }};
  var kappa = {{ kappa }};
  var xmax = {{ pmax }};
  var xbins_n = {{ xbins_n }};
  var ybins_n = {{ ybins_n }};
  var forecast_div_elm = "forecast_widgets_container";
</script>

<script type="text/javascript"
        src="{% static 'asset_simple/forecast.js' %}">
</script>

<div id = "forecast_widgets_container">
</div>
<script>
  forecast_instance = add_forecast_widget("forecast");
</script>
```