European Economic Review 61 (2013) 186–204

Contents lists available at SciVerse ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Learning, forecasting and optimizing: An experimental study

Te Bao [a,c], John Duffy [b,*], Cars Hommes [a]

[a] CeNDEF, School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018XE Amsterdam, The Netherlands
[b] Department of Economics, University of Pittsburgh 4901 Posvar Hall, Pittsburgh, PA 15260, USA
[c] IEEF, Faculty of Economics and Business, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 23 July 2012
Accepted 4 April 2013
Available online 16 April 2013

JEL classification:
C91
C92
D83
D84

Keywords:
Learning
Rational expectations
Optimization
Experimental economics
Bounded rationality

ABSTRACT

Rational Expectations (RE) models have two crucial dimensions: (i) agents on average correctly forecast future prices given all available information, and (ii) given expectations, agents solve optimization problems and these solutions in turn determine actual price realizations. Experimental tests of such models typically focus on only one of these two dimensions. In this paper we consider both forecasting and optimization decisions in an experimental cobweb economy. We report results from four main experimental treatments: (1) subjects form forecasts only, (2) subjects determine quantity only (solve an optimization problem), (3) they do both and (4) they are paired in teams and one member is assigned the forecasting role while the other is assigned the optimization task. All treatments converge to Rational Expectation Equilibrium (REE), but at different speeds. We observe that performance is the best in treatment 1 and the worst in Treatment 3. We further find that most subjects use adaptive rules to forecast prices. Given a price forecast, subjects are less likely to make conditionally optimal production decisions in Treatment 3 where the forecast is made by themselves, than in Treatment 4 where the forecast is made by the other member of their team, which suggests that “two heads are better than one” in term of the speed of finding the REE.

1. Introduction

Rational Expectations (RE) macroeconomic models have two crucial dimensions: (i) Rational agents on average correctly forecast future prices given all available information, that is, they do not make systematic forecast mistakes; (ii) Given agents' rational expectations, these same agents solve optimization problems that determine their consumption and/or production decisions, which then, via market clearing, determine the realizations of prices and wages the agents were seeking to forecast; these data are then used to update forecasts. Thus, RE systems are *self-referential*; beliefs affect outcomes and outcomes affect beliefs.

Testing rational expectation models with field data is problematic as agents' expectations are not generally observable and economists may disagree as to what constitutes the “true” model in which agents' expectations are formed. An alternative approach is to test rational expectations models in the laboratory where it is possible to control the model that determines economic data and to elicit and use agents' expectations of future variables in the determination of that same data. However, the self-referential nature of RE models makes it difficult to test these models in the laboratory.

* Corresponding author. Tel.: +1 412 648 1733; fax: +1 412 648 1793.
E-mail address: jduffy@pitt.edu (J. Duffy).

0014-2921/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.euroecorev.2013.04.003

As Sargent (2008) observes:

> "Laboratory experiments using macroeconomics are rarer than those using microeconomics...I suspect that the main reason for fewer experiments in macro than in micro is that the choices confronting artificial agents within even one of the simpler recursive competitive equilibria used in macroeconomics are very complicated relative to the settings with which experimentalists usually confront subjects."

Experimentalists seeking to test RE macroeconomic models have dealt with the complicated nature of these models by reducing the dimensionality of the problem that subjects face. Two approaches have been taken.

In a "learning to forecast experiment," (LtFE) – a design first proposed by Marimon and Sunder (1993) – subjects are asked to submit a forecast for a future economic variable (e.g., a price, inflation rate, foreign exchange rate, etc.), and they are rewarded solely on the basis of the ex-post accuracy of their forecast. Their forecast is used as an input by a computer program to determine each individual's optimal quantities as if the subjects themselves were capable of solving the optimization problem conditional on their forecast. The computer-determined quantities together with market clearing conditions then determine the actual price realizations (the object of the subjects' forecasts), and these realizations are then used to assess the accuracy of the subjects' forecasts. Subjects, however, are not necessarily made aware of how their forecasts affect outcomes; the mechanism by which subjects' forecasts determine the actual realizations of forecasted variables often amounts to a "black-box" process.

In a second, older experimental approach, known as a "learning to optimize experiment" (LtOE), subjects are asked to make economic decisions (to consume, invest, trade, produce, etc.) *directly*, without any elicitation of their forecasts of the relevant endogenous variables such as the market price, interest rate or wages. (e.g., Arifovic, 1996; Smith et al., 1988) Of course, such forecasts can be determined implicitly based on subjects' decisions or are sometimes determined separately via some market mechanism (e.g., a double auction or a call market) that is often external to the theory being tested.

Studies using the LtFE approach find mixed evidence as to whether subjects are able to learn a rational expectations equilibrium (REE) (see, e.g., Hommes, 2011 for a survey). In some instances, subjects learn a REE via some adaptive learning process while in other instances subjects behave as trend extrapolators resulting in persistent deviations or cycles around the rational expectations equilibrium (e.g., Anufriev and Hommes, 2012). Similarly, findings from LtOE studies have sometimes confirmed competitive equilibrium predictions and associated comparative statics predictions, but in other instances have generated outcomes that are at odds with RE model predictions, for instance, non-rational bubbles, excess volatility, etc.

In this paper we compare the LtFE and LtOE approaches in a common, economic decision-making task. Importantly, we also consider how behavior improves or deteriorates if we combine these two approaches. Our combined LtFE and LtOE design gets at the heart of the belief–outcome interaction that is the signature property of rational expectations models. We ask if convergence to the REE and efficiency are affected when subjects are asked to play both roles as forecaster and optimizer or if specialization of tasks by individuals alone (as in LtFE and LtOE designs) or within two-agent teams leads to a significant improvement in performance. One aim of this research is to assess whether the results obtained in the LtFE literature are robust when the optimization task is performed by an individual rather than by a computer program. Moreover, our novel team specialization treatment has a very natural, real-world interpretation: Organizational investors such as investment banks and pension funds usually employ both professional forecasters (researchers and economists) and production managers or traders. This type of team specialization set-up has not been previously explored in the laboratory.

The experimental environment we study is a simple, N-firm cobweb model economy—a negative expectation feedback system. This kind of feedback system arises naturally in commodity markets that were the inspiration for Ezekiel's (1938) development of the cobweb model. Furthermore, Muth (1961) proposed rational expectations in the context of this very same negative feedback cobweb model. Prior research indicates that under a LtFE design, market prices will converge very quickly to the RE equilibrium in this environment. In addition to LtFE, we consider three additional treatments where subjects must submit their production decision directly without a forecast (LtOE), or together with a forecast, or subjects are paired in teams and one team member submits a forecast which the other team member can use to determine a production decision.

We find a tendency for the market price to converge to the REE price in all four treatments. Thus, the stabilizing effect of a negative feedback market is a robust feature of our experiment. However, when the speed of convergence is compared across treatments, we find that the market price converges most quickly and reliably when subjects only make price forecasts as in the computer-aided LtFE design. There is not much difference in performance between the treatment where subjects only make production decisions (LtOE) and the treatment where they form teams that specialize in one of the two tasks. However, the market price and quantity fluctuate the most and are the slowest to converge when subjects are required to perform both forecasting and production decision-making (optimizing) tasks. Our findings have important implications not only for the design of experiments, but more importantly for how we might think about the representative agent firm: should it be viewed as an individual actor (e.g., the C.E.O.) or is it better to think of the representative firm as consisting of teams of individuals specialized in various tasks, such as forecasting and production? Further, our decomposition of the forecasting and optimization tasks suggests that bounded rationality with respect to *optimization* decisions appears to be as important a consideration in the learning of rational expectations equilibria as is bounded rationality in expectation formation. Nevertheless, most research on learning in macroeconomics (e.g., Evans and Honkapohja, 2001) has focused only on bounded rationality with respect to expectation formation and not with respect to optimizing behavior.

The rest of the paper is organized as follows: Section 2 discusses the related literature. Section 3 describes our experimental design. Section 4 presents the experimental results. Finally, Section 5 concludes.

2. Related Literature

Our work is related to prior studies using either the LtFE or LtOE designs. Smith et al. (1988), Lim et al. (1994), Arifovic (1996), Lei et al. (2001), Noussair et al. (2007) and Crockett and Duffy (2012) are some examples of LtOE studies. Adam (2007), Marimon et al. (1993), Marimon and Sunder (1993, 1994, 1995), Hommes et al. (2005, 2007), Heemeijer et al. (2009) and Bao et al. (2012) are some representative studies using the LtFE design.

As we also have a treatment where subjects participate as members of teams, our experiment is related to the literature on the comparison of group and individual decisions. In the experimental macroeconomics literature, Blinder and Morgan (2005) and Lombardelli et al. (2005) show that monetary policy decisions made by groups of subjects acting a central bankers are not slower than the policy decisions made by individuals playing the same role as central banker, and that the group decisions are generally better (in the sense of minimizing a loss function). There is a parallel literature in experimental game theory on individual versus group decisions. The evidence from that literature is mixed on whether groups are more "rational" or self-interested than individuals. For instance, Kocher and Sutter (2005) find that groups learn faster, and can beat individuals in play of a "beauty-contest" game. Cooper and Kagel (2005) report that groups act more strategically than individuals in the context of a signaling game. Bornstein and Yaniv (1998) find that groups offer less and accept less in the "ultimatum game" relative to individuals. However, Cox (2002) shows that there is no significant difference between group and individual decisions in the "trust game." Cason and Mui (1997) find that groups offer more in "dictator games" than individuals. In all of these group-versus-individual-studies, group members are asked to perform/participate in the *same* kind of task and the decision of the group is usually the average or majority choice of all group members. By contrast, our group (team) treatment involves *specialization* of tasks between the two group members, who share a common interest in maximizing their joint payoff; we are not aware of any prior group–versus–individual study with this type of specialization of tasks among the group members.

Our work is also related to experiments studying Cournot oligopoly models. Offerman et al. (2002) demonstrate that providing subjects, in the role of firms, with different information about the behavior of other firms (subjects), e.g., information about the sum of other firms' quantity choices only, about individual firm's quantity choices only or about individual firm's quantity choices and profits, can lead to the adoption of different learning rules and market evolution toward different equilibria (Walrasian, Collusive and Cournot-Nash). In our experiment, subjects playing the role of firms have no information about other firms' quantity choices or profits. They also have no information about the relationship between the market price and total output. As the optimal quantity decision requires them to set price equal to marginal cost, the rational expectations equilibrium in this Cournot market is the same as the Walrasian outcome. Huck et al. (1999) vary the information available to subjects from full information about the market including others' decisions and profits and their own decision and profit to information only about their own decision and profit. They report that none of their information treatments generate successful collusion and that information that encourages "imitate the best" learning leads to a Walrasian outcome, which confirms the prediction of Vega-Redondo (1997). Their "NOIN" treatment, where subjects have no information about the behavior of other subjects (firms), is similar to the information that we provide to our subjects except that their subjects know the number of firms in the market, but do not have a payoff table informing them of their profit as a function of the chosen quantity. Their NOIN treatment generates an outcome that is very close to the Walrasian outcome and that is why we chose this informational structure for our experiment. However, as their environment involves a constant marginal cost, the optimal quantity given a price prediction is piecewise linear and generates no steady state. It is therefore not possible to test convergence to REE using their experimental design, and hence our use of a different design.

3. Experimental design

3.1. Theoretical model

The model behind our experiment is one of demand and supply for a single, non-storable good. Denote by D the non-negative and monotonically decreasing aggregate demand function for this good and by $S_{h,t}$, the non-negative supply function for the good by firm h, derived from expected profit maximization at period t. Let $p^e_{h,t}$ be the price forecast made by firm h in period t. The supply function may be rewritten as $S(p^e_{h,t})$. We assume that all firms have the same supply function. Subjects play only the role of firms (suppliers) in the experiment; demand is exogenously given (as described below).

The market price is determined by the market clearing condition for a cobweb economy, which is given by

$$p_t = D^{-1}\left(\sum_h S_{h,t}\right) + \epsilon_t, \tag{1}$$

where $\epsilon_t \sim N(0,1)$ is the realization of an i.i.d. price shock in period t. In the experiment, the sequence of realizations of the noise term is the same across all treatments.

We assume there are H firms (suppliers), differing only in the way they form expectations. We use a linear demand function $D(p_t) = a - bp_t$, where $a = 63, b = 21/20$. We assume that each firm has a cost function $c(q) = Hq^2/2$. The expected profit of firm h, $\pi^e_{h,t}$, is given by

$$\pi^e_{h,t} = p^e_{h,t} q_{h,t} - c(q_{h,t}). \tag{2}$$

Solving the profit maximization problem yields the optimal supply function for each firm: $S^*(p^e_{h,t}) = p^e_{h,t}/H$. If every firm makes supply decisions optimally, the total supply on the market will coincide with the mean price forecast, $\sum_h S^*(p^e_{h,t}) = \overline{p^e_t}$. Substituting this optimal market supply into the market clearing condition (Eq. (1)), we have that

$$p_t = \max\left\{\frac{20}{21}(63 - \overline{p^e_t}) + \epsilon_t,\ 0\right\}. \tag{3}$$

Imposing the RE assumption ($p^* = E[\frac{20}{21}(63 - p^*) + \epsilon_t]$), and noting that the expected value of the noise term is zero, we find the rational expectations equilibrium (REE) price, $p^* = 30.73$. The optimal supply in this REE is 5.12, and the profit for each firm is 78.70.

Subjects were not informed of the precise demand function as detailed in this section nor were they informed of the total quantity supplied (the quantity decisions of the other $H-1$ firms in their market). However, they were told that market demand was decreasing in the market price and that the market price was determined by market clearing, i.e., that supply equals demand—see the Instructions in Appendix A for the specific details.

3.2. Treatments

Our experiment consists of four main treatments that differ in the tasks assigned to participants and in the payoff scheme. A fifth treatment explores whether a change in the payoff scheme matters for observed differences between some of the treatments. Sample experimental instructions are provided in Appendix A. Subjects in our experiment play the role of firms only, deciding on price forecasts or on optimal amounts of production or both.

1. Treatment 1: The LtFE treatment. In this treatment, subjects (firms) only make price forecasts (or “predictions”). Given each firm's price forecast, their (conditionally) optimal production decision is calculated for them by a computer program. Aggregating the H firms’ conditionally optimal supply decisions and equating that aggregate supply with exogenous aggregate market demand yields the actual market price. Each subject is paid according to the accuracy of his price forecast, which is a function of the difference between his forecast price and the actual market price. Each subject knows the prior history of the market price they are attempting to forecast which is standard in the LtFE literature and the history of their own past forecasts and payoffs. Each subject was given a payoff table that showed their payoff from the forecasting task for different “prediction errors” (See Appendix D, “Payoff Table for Forecasters”).
2. Treatment 2: The LtOE treatment. In this treatment, subjects (firms) make quantity (or “production”) decisions and there is no computer assistance. Each subject knows the history of the market price, his own prior decisions and profits. Each subject makes a quantity decision only; there is no elicitation of a subject's price forecast. The market price is determined by the production decisions submitted by all H firms in the market as equated with exogenous market demand. Each subject is paid according to the profit his firm makes each period. Subjects are given a table showing their potential payoff (profit) for different combinations of the market price and the subject's own production (optimization) decision (See Appendix D, “Payoff Table for Production Managers”).
3. Treatment 3: The LtFE+LtOE individual treatment. In this treatment, each subject plays the role of both forecaster and production manager. Each subject knows the history of the market price and his prior decisions and profits. Each subject makes both a price forecast and a quantity decision. The market price is again determined by the production decisions submitted by all firms in the market as equated with exogenous market demand. Subjects are paid according to an equal weighted linear combination of the payoff functions used in the LtFE and LtOE treatments. Each subject can read his potential payoffs for the forecasting task from the payoff table for forecasters and his potential payoffs from the production (optimization) task from the payoff for quantity decisions (same tables as in Treatments 1 and 2, respectively).
4. Treatment 4: The LtFE+LtOE team treatment. In this treatment, there is a forecaster and a production manager in each two-agent team. The forecaster knows the history of market prices and the production manager knows the history of his own production decisions and profits. The market price is determined by the production decisions of all firms in the market in combination with exogenous market demand. Each subject is paid according to an equal weighted linear combination of the payoff functions used in the LtFE and LtOE treatments, exactly as in Treatment 3. Subjects can read the potential payoffs for the forecasting task from the payoff table for forecasters and the potential payoffs for the production task from the payoff for quantity decisions (same tables as in Treatments 1 and 2, respectively).
5. Treatment 5: The LtFE treatment where the subjects receive payoffs from the implied quantity decision, instead of from their forecast accuracy. This treatment was motivated by the comment by a referee of this paper who observed that differences in both the nature of the task (forecasting/optimizing) or the payoff functions. Eqs. (4) and (5) (shown below)

could account for the observed differences between Treatments 1 and 2. In particular, errors made by subjects are punished more severely under the payoff function (4) used for Treatment 1 than under the payoff function (5) used for Treatment 2.[1] Therefore, this treatment addresses whether differences between the LtFE and LtOE treatments arise from the different payoff structures used in the LtFE and LtOE designs as opposed to the different tasks (forecasting or optimizing). First, each firm forms a price forecast as in Treatment 1. Given this forecast, their (conditionally) optimal production decision is calculated for them by the computer program. Aggregating the H firms' conditionally optimal supply decisions and equating that aggregate supply with exogenous aggregate market demand results in an actual market price. Each subject is paid according to the *profit* his firm makes each period as in Treatment 2. Subjects are given a table showing their potential payoff (profit) for different combinations of the market price and the subject's own price forecast.

In all treatments, we restricted price forecasts to be non-negative and less than 100. Note that price forecasts in excess of 60 would result in negative prices, and thus our upper bound on price predictions is not very restrictive. The quantity decision was also restricted to be non-negative, and we set 20 as the upper bound for the quantity decision as the payoff for the production manager would be negative if he chose to produce more than 20 units when the price is 0.

3.3. Number of observations

We report results from 9 experimental sessions that were conducted using the CREED laboratory at the University of Amsterdam on April 27–29, May 3, 2011 and November 28, 2012. A total of 198 subjects participated in the 9 sessions of this experiment. These subjects are mainly bachelor and master students at the University of Amsterdam. No subject participated in more than one session, and all subjects in the same session faced the same treatment condition. Each session involved multiple groups of $N=6$ or $N=12$ participants who interacted with one another for 50 periods in one of our five treatments, that is, we adopt a "between subjects" design. We refer to each independent observation, involving $N=6$ or 12 subjects interacting together for 50 periods under the same Treatment conditions as a "market." Among the 9 sessions, 1 session is allocated to Treatment 1, 2 are allocated to each of treatment 2 and 3, 3 sessions are allocated to Treatment 4 and 1 session is allocated to Treatment 5. We recruited up to 30 subjects for each session. Depending on the show up rate, subjects were divided up to form 3 or 4 markets in Treatments 1, 2, 3 and 5, and 1 or 2 markets in Treatment 4. The number of observations (markets) is about 6 for each of Treatments 2, 3 and 4. We have fewer observations for Treatment 1 because a very similar experiment was previously conducted by Hommes et al. (2007) and Heemeijer et al. (2009) and the results from all 4 markets of our Treatment 1 are very similar to the results obtained in these prior studies. Treatment 5 has just 3 markets as we were only seeking to understand whether a fairly neutral change in the payoff structure affected our findings. A summary of the number of markets (observations) and the number of participants per market for each of our treatments is given in Table 1.

Notice that in Treatments 1, 2, 3 and 5 we always had 6 subjects (or firms) per market, while in our team Treatment 4 we had 12 subjects per market so that each of the 6 "firms" consisted of a pair of players (a "team") who remained matched together for all 50 periods of the market. The duration of each session depended on the treatment. Treatments 1 and 5 involving forecasting only took around 70 min to complete. Treatment 2 involving quantity choices only (optimization) took around 90 min to complete. Finally, Treatments 3 and 4 which involved both forecasting and optimizing took about 120 min to complete.

Prior to the first period of each market, we asked subjects to answer some comprehension ("control") questions designed to test their understanding of the written instructions. Subjects were not allowed to begin participating until they had correctly answered all of these questions. The instructions and control questions are given in Appendix A. The payoffs earned by subjects were between 15 and 25 euros (Table 3 provides further payoff details).

3.4. Computer interface

Fig. 1 provides an illustration of the computer interface that subjects saw in the experiment, specifically for Treatment 4. The screen was divided into 3 mini pages. In the top mini page, subjects were prompted to submit their decisions, i.e., their price forecast or their quantity production choice. In the bottom left mini page they saw a graph plotting past market prices (the "Real Price") and, if they were a forecaster, they also saw their past price forecast history ("Your Prediction"). Finally, in the bottom right mini page they saw a table reporting the history of realized (market) prices, as well as their own prior decisions and their period and cumulative payoffs.

The top panel of Fig. 1 shows the computer interface that forecasters saw in Treatment 4. The computer interface the forecasters saw in Treatments 1 and 5 is very similar to the one shown for forecasters in Treatment 4, except that the

[1] The payoff functions that are used for LtFE and LtOE type experiments are generally of the type given by Eqs. (4) and (5), respectively, and that is why we chose these functional forms. We thank the referee for pointing out the potential confound in using different payoff functions in making comparisons between these two treatments.

Table 1
Characteristics of the experimental design.

Treatment number	Number of firms per market	Number of participants per market	Total number of markets (Observations)	Total number of participants
1	6	6	4	24
2	6	6	7	42
3	6	6	7	42
4	6	12	6	72
5	6	6	3	18

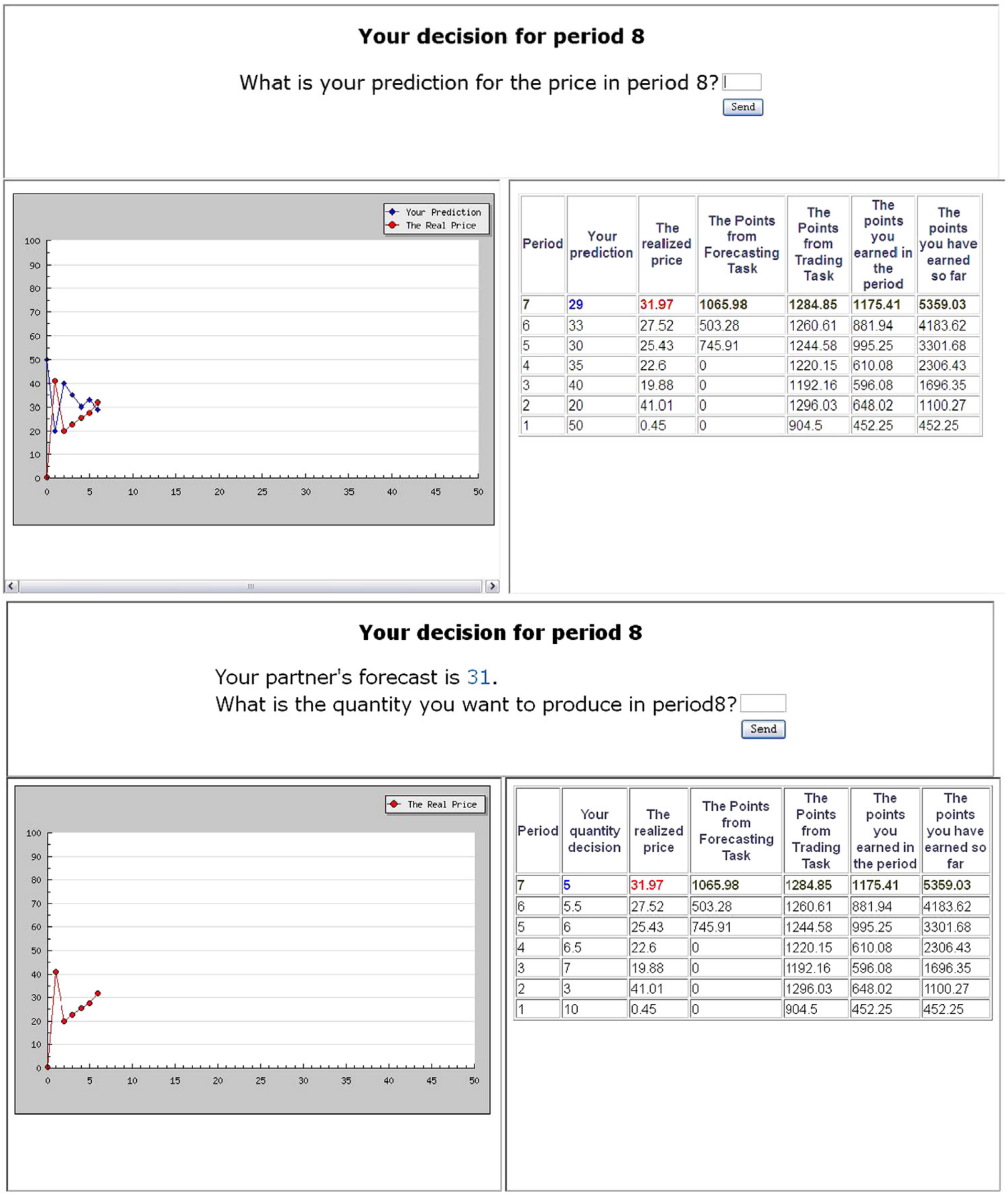

Fig. 1. The computer interface for forecasters (top) and production managers (bottom).

history of past performance (points earned) was shown only for the forecasting task and not for the production task as in Treatment 4.

The bottom panel of Fig. 1 shows the computer interface that production managers saw in Treatment 4. At the start of each period these production managers saw a notice on their screen: “we wait for your partner to give a forecast.” Once the forecaster/team partner had submitted his/her forecast, the production manager was informed of this forecast (as shown in the bottom panel of Fig. 1) and the production manager was then asked to enter a quantity decision for the team.

The computer interface that subjects see in Treatment 2 is very similar to that shown in the bottom panel of Fig. 1 except that there is no waiting phase and the history of past performance is only for the optimization (production) task instead of for both the forecasting and production tasks as in Treatment 4. The computer interface in Treatment 3 is also similar to the one shown in Fig. 1 except that there is no waiting phase and the same subject is asked to first submit a price forecast and then to submit a quantity decision. The history of past performance for Treatment 3 is the same as for Treatment 4 as the payoff functions are the same in these two treatments.

We note that there were no time constraints on decision-making in any of our treatments. Nevertheless, we recorded data on the time it took subjects to make certain decisions, as discussed later. The market price was not determined until all N subjects had submitted their price forecasts and/or quantity production decisions; subjects were instructed to wait until all decisions for the period had been finalized. At the end of each period, the computer program calculated the market price (in the manner described above) and reported this market price back to subjects along with their earnings from their forecasting or production decisions (or both) for the period. The historical information on the decisions screens was refreshed and if the last (50th) period had not yet been played, a new period would then begin. Each period took no more than 5 min to complete (and was often much faster than that).

3.5. Payoffs

Subjects earned points during the experiment that were converted into euros at the end of the experiment at a known and fixed rate. The payoff function for forecasters in points (in all but Treatment 5) is a decreasing function of their prediction error, and was given by

$$\text{Payoff for Forecasting Task for Firm } h = \max\left\{1300 - \frac{1300}{49}(p_t - p^e_{h,t})^2, 0\right\}. \tag{4}$$

Notice that subjects earn 0 if their price forecast error is greater than 7, and they earn a maximum of 1300 for a perfect forecast.

The payoff function for the production (optimization) task (in points) was given by

$$\text{Payoff from the Production Task for Firm } h = p_t q_{h,t} - 3(q_{h,t})^2 + 1200. \tag{5}$$

Notice that subjects get a baseline 'salary' of 1200 points plus the actual profit earned by their firm, which depends on the market determined price, p_t and on the quantity, $q_{h,t}$, chosen by their firm. A firm's profit can be negative, so a subject's payoff can be smaller than 1200. However, our set-up implies that the maximum possible loss (the absolute value of negative profit) is 1200 (which is the loss a firm will make when the market price is 0 and it is producing 20 units of goods), so that each subject's total payoff can never be negative. As the profit for the firm when the market price equals the REE price is about 80, the maximum payoff earned by a subject as a forecaster or as a production manager is approximately the same, at around 1300 points per period.

Subjects in Treatment 1 earn the payoff from the forecasting task only. Subjects in Treatment 2 earn the payoff from the production task only. Subjects in Treatments 3 and 4 each earn the equal weighted average of the payoffs from the forecasting and production tasks. Subjects in Treatment 5 perform the forecasting task but are paid according to the payoff table used for the production task of Treatment 2. The payoff function for the forecasting task in Treatment 5 is given by

$$\text{Payoff for Forecasting Task for Firm } h \text{ in Treatment } 5 = p_t \frac{p^e_{h,t}}{6} - 3\left(\frac{p^e_{h,t}}{6}\right)^2 + 1200. \tag{6}$$

Note that (6) is similar to (5) except that $q_{h,t}$ is replaced by the conditionally optimal quantity given firm h's forecast, $p^e_{h,t}/6$.

These payoff functions were all carefully explained to subjects in the written instructions and were presented to them as payoff tables (see Appendices A and D). At the end of the experiment, subjects were paid 1 euro for each 2600 points they earned in all 50 periods of all treatments of the experiment and this conversion rate was known to subjects in advance.

4. Experimental results

4.1. Aggregate market price

Fig. 2 plots the average market prices in each treatment against the REE price, $p^* = 30.73$. This figure reveals that the average price in all five treatments gets very close to the REE price, especially in the later periods of the experiment. Thus, the general tendency for a negative feedback system to converge to REE is not greatly affected by the type of task (forecasting or optimizing or both) that is assigned to market participants. However, the adjustment toward the REE at the beginning of the experiment is (as we document below) fastest in Treatment 1 and slowest in Treatment 3. Treatment 5 also achieves very fast convergence to the REE in the initial periods however there were also some fluctuations in the middle of that treatment that were caused by some subjects "experimenting" with very large or low predictions. The results suggest that the nature of the task (forecasting or optimizing) seems to be the force that determines the speed of the convergence to a neighborhood of the REE in the initial periods, while the payoff incentive structure (forecast accuracy or profit maximization) seems to be the force that determines whether prices remain steadily close to the REE once convergence to

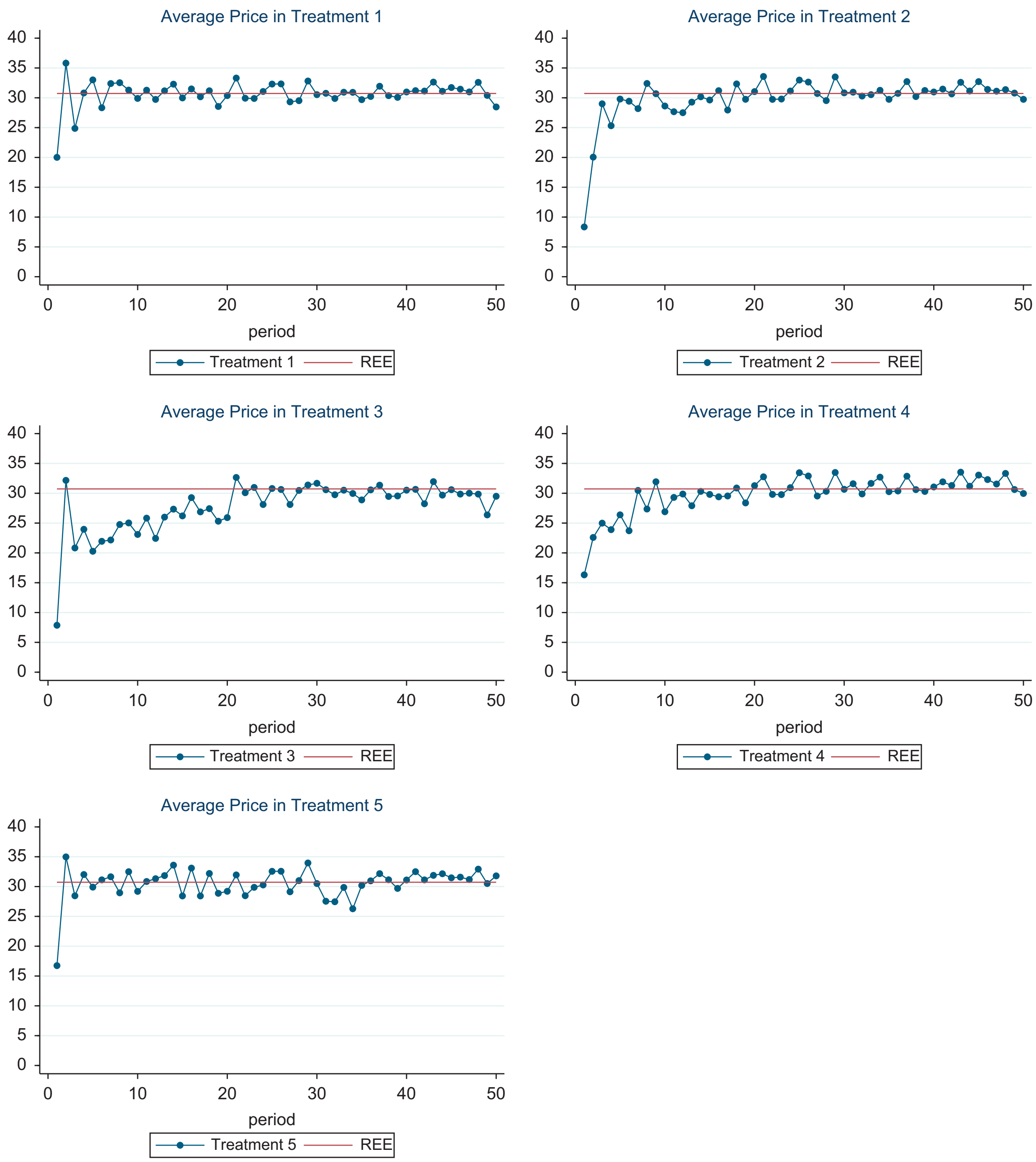

Fig. 2. The average market price against the REE price in each of the five treatments.

the REE has been achieved. Under the forecasting task, subjects appear to learn the REE more quickly than under the quantity decision task, but with the relatively flat incentive structure of the quantity decision task they do not lose very much by deviating from the REE after achieving some degree of convergence, which leads some of them to behave more erratically under that profit-maximizing payoff incentive structure. The volatility of the market price is also smallest in Treatment 1, and largest in Treatment 3.

As a first check on whether prices are, in fact, converging to the RE prediction, we declare convergence to have occurred in the first period for which the difference between the market price and the REE price is less than 5 and stays below 5 forever after that period. Using this simple criterion, we count the number of periods required for convergence across our different treatments, as reported in Table 2. If there is no convergence according to our criterion, as is the case for 5 markets in Treatment 3, then we count the number of periods to convergence as the full sample size of 50 periods. Comparing these time–to–convergence numbers, we observe that the market price converges faster in Treatment 1 than in each of the treatments 2, 3 and 4 (the difference is significant at the 5% level according to a two sided Wilcoxon Mann–Whitney test

Table 2
The number of periods to convergence for each market.

Treatment	Market	Number of periods to convergence
Treatment 1	Market 1	3
	Market 2	3
	Market 3	4
	Market 4	1
	Mean	2.75
	Median	3
Treatment 2	Market 1	17
	Market 2	33
	Market 3	13
	Market 4	12
	Market 5	11
	Market 6	4
	Market 7	28
	Mean	14.43
	Median	13
Treatment 3	Market 1	50
	Market 2	50
	Market 3	35
	Market 4	3
	Market 5	50
	Market 6	50
	Market 7	50
	Mean	42.29
	Median	50
Treatment 4	Market 1	36
	Market 2	10
	Market 3	13
	Market 4	25
	Market 5	6
	Market 6	10
	Mean	10.67
	Median	10
Treatment 5	Market 1	33
	Market 2	4
	Market 3	31
	Mean	22.67
	Median	31

using the independent market observations for each treatment), and there is no significant difference between Treatments 1 and 5.[2] We further observe that convergence is faster in Treatments 2, 4 and 5 than in Treatment 3 (the difference is significant at the 5% level according to a Wilcoxon Mann–Whitney test). Finally, the difference between each pair of Treatments 2, 4 and 5 is not significant at 5% level according to a Wilcoxon Mann–Whitney test.

We can also test for convergence econometrically using a method suggested by Duffy (2012). For each market j, the following linear equation is estimated:

$$p_{j,t} = \lambda_j p_{j,t-1} + \mu_j + \epsilon_{j,t} \tag{7}$$

The results of this estimation exercise are reported in Appendix B. We note first that all of the estimated μs are significantly different from 0 at the 5% level of significance, and all the estimated λs are significantly different from 0 at the 5% level except markets 23 (the third market in Treatment 2), 34 (the fourth market in Treatment 3) 42 (the second market in Treatment 4) and 53 (the third market in Treatment 5). We also checked for evidence of serially correlated errors. For our estimation, the relevant upper bound of the Durbin–Watson statistic, dU, $(n = 50, k' = 2)$ is 1.445. We found that for each market, the estimated Durbin-Watson statistics were always greater than that upper bound, which implies that we cannot reject the null hypothesis of no first order serial correlation in the error terms. The estimated linear equation is stable if $|\lambda|$ is smaller than 1.[3] and has a long-run equilibrium level $\hat{\mu}_j/(1-\hat{\lambda}_j)$. For each market j, we declare that *weak convergence* obtains

[2] But this may be due to the small number of observations in these treatments: the number of observations is only 4 for Treatment 1, and 3 for Treatment 5.

[3] As all the estimated λs are positive, we just need to check whether $\lambda \geq 1$ is rejected. This statement is equivalent to the claim that the price dynamics are stationary.

if we can reject $\hat{\lambda}_j \geq 1$ at the 5% level, and we say that *strong convergence* obtains if we cannot reject $\hat{\mu}_j/(1-\hat{\lambda}_j) = 30.73$ (the REE value) at the 5% level (using a Wald test). Summarizing the estimation results (as reported in Appendix B), we have:

1. All markets in all five treatments satisfy *weak convergence*.
2. All markets in Treatments 1 and 2, 5 satisfy *strong convergence*. All but one market in Treatment 4 satisfies strong convergence. The equilibrium price in the one market of Treatment 4 that does not satisfy strong convergence is not very different from the REE ($\hat{\mu}_j/(1-\hat{\lambda}_j) = 32.08$). Only 2 out of 7 markets in Treatment 3 satisfy strong convergence.

We see a large difference between Treatment 3 and the other three treatments. The difference between Treatments 3 and 4 in particular suggests that teamwork and specialization may help participants to find the REE.

4.2. Individual-level decisions

4.2.1. Distribution of decisions

We have seen that the aggregate market price gets close to the REE price in many markets. It is of interest to consider whether decisions at the *individual* level are also consistent with RE predictions. The empirical cumulative distribution function (CDF) of individual price forecasts and optimization (quantity-choice) decisions is shown in Fig. 3 using pooled data from all markets of the relevant treatments (Treatments 1, 2, 4 and 5 for the price forecasts and Treatments 2, 3 and 4 for the quantity choices). Under rational expectations, the CDF should be a step function switching from 0 to 100% at the REE price (re = 30.73) or REE quantity (qre = 5.12).

Fig. 3 reveals that there is some heterogeneity in individual decisions across treatments with the largest departures from RE predictions occurring in Treatment 3, a finding that is consistent with our findings using the more aggregate measures of market prices and quantities.

Using the distribution of individual forecasts for the three treatments involving forecasting, we perform a one-sample Kolmogorov–Smirnov test of whether the distribution of individual forecasts is significantly different from the RE prediction, $p^* = 30.73$ for all 50 periods, the first 25 periods, or the second 25 periods. For all three samples, we can reject the null hypothesis of no difference in price forecasts for all five treatments at the 5% level. The top panel of Fig. 3 suggests that the distribution of individual forecasts is similar in Treatments 1, 4 and 5, while Treatment 3 looks very different. For confirmation we performed a two-sample Kolmogorov–Smirnov test on whether the distribution of individual forecasts is the same between each possible pairing of these three treatments and for all three samples (all 50, first 25 or second 25 periods) and we find that each treatment is significantly different from each of the other treatments for all samples (at the 5% level). Indeed, the ordering is such that Treatment 1 is closest to the RE price prediction, Treatment 3 is furthest and Treatments 4 and 5 are intermediate.

For the distribution of individual quantity decisions, we also perform a one-sample Kolmogorov–Smirnov test on whether the distribution of individual quantity decisions is significantly different from the RE prediction that all firms produce 5.12 units for all 50 periods, the first 25 periods, or the second 25 periods. For all three samples, we can reject the null hypothesis of no difference in quantity choices for all three treatments at the 5% level. The lower panel of Fig. 3 suggests that the distribution of individual quantity decisions is similar in Treatments 2 and 4, while Treatment 3 looks very different. We again performed a two-sample Kolmogorov–Smirnov test on whether the distribution of individual quantity decisions is the same between each possible pairing of the three treatments. The tests indicate that there is no significant difference in the distribution of quantity decisions between Treatments 2 and 4 in the sample involving all 50 periods or the second 25 periods, but not in the first 25 periods, and there is a significant difference between Treatment 3 and the other two treatments (at the 5% level) in all the three samples. In particular, there is much greater heterogeneity in the quantity decisions of Treatment 3 as compared with either Treatments 2 or 4.

4.2.2. Time taken to make decisions

We also collected data on the time it took subjects to make their decision(s). Such data can be useful in understanding the cognitive difficulty of decision-making. In particular, Rubinstein (2007) provides evidence that choices requiring greater cognitive activity are associated with longer decision response times. While there was no decision time limit in our experiment (subjects could take as much time as they wished for each decision), the computer program that implements our experiment started counting (in seconds) when a subject first entered each new period, and stopped counting when he or she submitted his or her decision(s). Fig. 4 plots the empirical CDF of the time taken by subjects in each period of Treatments 1, 2, 3 and 5.[4] As subjects submit their forecasts and quantity decisions together in Treatment 3, the decision-time data for Treatment 3 is the total time taken for both the forecasting and optimizing tasks.

Fig. 4 clearly reveals that subjects take less time to make their decisions in Treatments 1, 2 and 5 as compared with Treatment 3. The average decision time per period is 19.83 s in Treatment 1, 21.16 s in Treatment 2, 33.99 s in Treatment 3 and 18.20 s in Treatment 5. The difference between either Treatment 1 or 2 and Treatment 3 is significant at the 5% level

[4] There was a technical problem with decision-time capture in Treatment 4, and as a consequence we cannot construct precise decision-time data for that treatment.

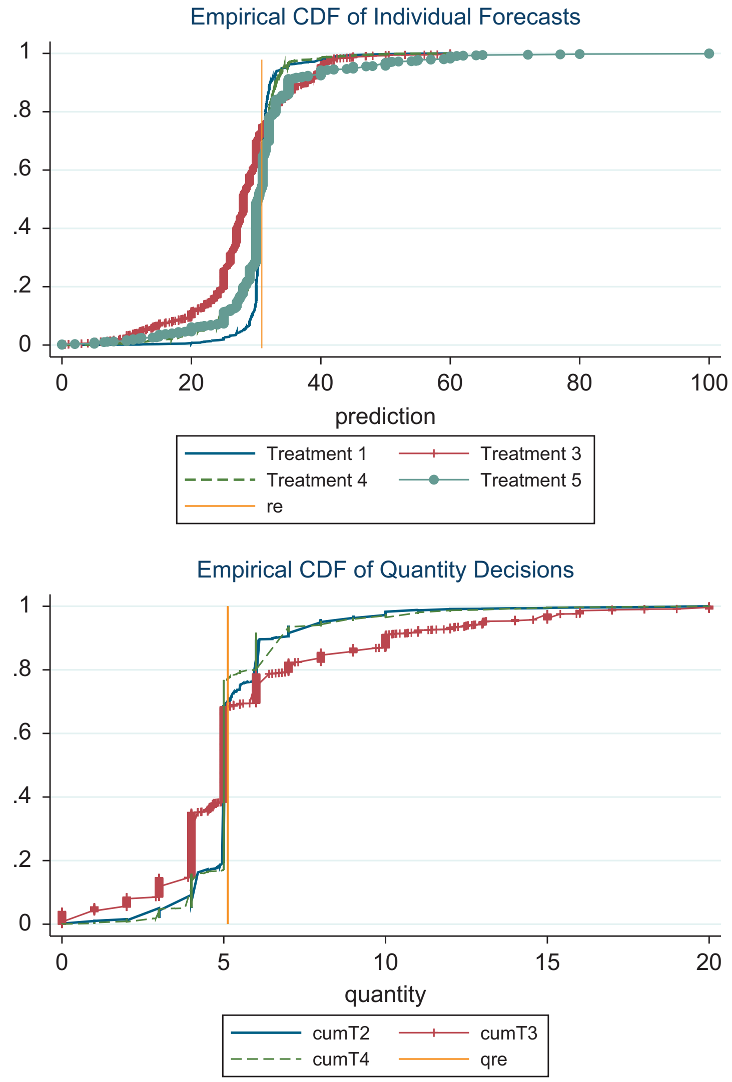

Fig. 3. The empirical cdf of individual price forecasts (top) and quantity decisions (bottom) over all 50 periods. The REE price and quantity are denoted by "re" and "qre" respectively, and are represented by vertical lines in the graphs.

according to a Wilcoxon Mann–Whitney test. The difference between Treatments 1 and 2 is not significant at the 5% level. The difference in time taken in Treatment 5 and either of Treatments 1 and 2 is significant at the 5% level according to a Wilcoxon Mann–Whitney test, which provides additional support for the notion that the forecasting task is viewed by subjects as being simpler than the optimizing task. These results suggest that subjects did not find the optimization task to be more cognitively difficult than the forecasting task but that making two decisions, as in Treatment 3, is indeed more cognitively challenging than making a single decision as in Treatments 1 and 2. The average decision time taken in Treatment 3 is, however, less than twice the average time taken in Treatments 1 and 2, and that difference is also significant at 5% level according to a Wilcoxon Mann–Whitney test.

4.3. *Earnings efficiency*

We compare subjects' earnings in the experiment to the hypothetical case where all subjects play according to the REE predictions in all 50 periods. Subjects can earn 1273.47 points per period for the forecasting task when they play according to REE because their prediction errors only come from the noise term ϵ_t, which means they earn 0.4898 Euro each period, and 24.49 Euros for all 50 periods. The profits they can earn for the production task is 1278.7 points per period when they play according to the REE, which means they earn 0.4918 Euro per period, and 24.59 Euros for 50 periods. We use the ratio of actual to hypothetical REE payoffs as a measure of efficiency. This measure can be greater than 100 percent in treatments with production decisions, because subjects can earn more by producing a little less than the REE prediction. These efficiency ratios, as reported in Table 3, are generally very high (exceeding 80%) in all five treatments. The ranking of average efficiency over all 50 periods is: Treatment 2 > Treatment 5 > Treatment 4 > Treatment 1 > Treatment 3, while the ranking for the last 25 periods is: Treatment 2 > Treatment 5 > Treatment 1 > Treatment 4 > Treatment 3. Only the difference between efficiency in Treatment 2 and each of the other three treatments is significant at the 5% level according to a

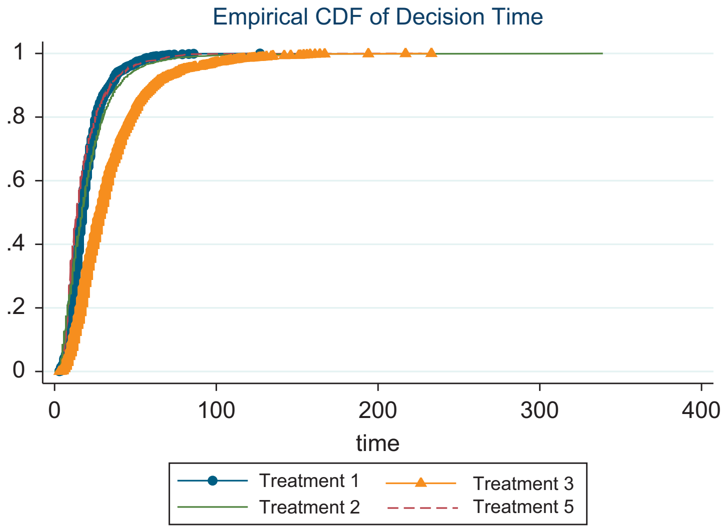

Fig. 4. The empirical cdf of the time taken to complete decision tasks in treatments 1, 2, 3 and 5. The unit of time is seconds, as measured on the horizontal axis.

Table 3
Average earnings (in euros) and efficiency for each market.

Treatment	Market	Period 1–50		Period 1–25		Period 26–50	
		Average earnings	Efficiency (%)	Average earnings	Efficiency (%)	Average earnings	Efficiency (%)
Treatment 1	Market 1	20.44	83.46	8.45	68.99	11.99	97.94
	Market 2	21.57	88.06	9.47	77.37	12.09	98.76
	Market 3	21.50	87.79	9.59	78.28	11.91	97.30
	Market 4	21.83	89.15	10.50	85.78	11.33	92.53
	Average	21.34	87.12	9.50	77.60	11.83	96.63
Treatment 2	Market 1	24.45	99.43	11.64	94.70	12.81	104.16
	Market 2	23.98	97.53	11.73	95.43	12.25	99.64
	Market 3	23.95	97.40	12.19	99.18	11.76	95.61
	Market 4	24.47	99.50	11.90	96.81	12.56	102.19
	Market 5	24.43	99.36	12.03	97.85	12.40	100.88
	Market 6	24.33	98.96	12.09	98.35	12.24	99.56
	Market 7	24.25	98.62	12.14	98.71	12.11	98.53
	Average	24.27	98.69	11.96	97.29	12.30	100.08
Treatment 3	Market 1	22.10	90.06	9.68	78.89	12.42	101.22
	Market 2	18.57	75.66	9.42	76.76	9.15	74.56
	Market 3	20.63	84.08	7.08	57.67	13.56	110.49
	Market 4	21.18	86.32	10.53	85.83	10.65	86.82
	Market 5	19.12	77.90	9.06	73.81	10.06	81.99
	Market 6	22.78	92.84	10.93	89.09	11.85	96.59
	Market 7	19.27	78.51	8.39	68.39	10.88	88.64
	Average	20.52	83.62	9.30	75.78	11.22	91.47
Treatment 4	Market 1	22.10	90.06	10.07	82.09	12.03	98.03
	Market 2	21.80	88.83	10.14	82.68	11.66	94.99
	Market 3	21.08	85.91	9.36	76.28	11.72	95.55
	Market 4	20.60	83.94	9.16	74.61	11.44	93.28
	Market 5	22.32	90.94	10.09	82.24	12.23	99.64
	Market 6	22.13	90.19	10.65	86.76	11.49	93.62
	Average	21.67	88.31	9.91	80.78	11.76	95.85
Treatment 5	Market 1	24.41	99.29	12.17	98.93	12.26	99.67
	Market 2	24.49	99.59	12.20	99.16	12.30	100.01
	Market 3	24.36	99.05	12.18	99.03	12.19	99.07
	Average	24.32	99.31	12.19	99.08	12.25	99.56

Wilcoxon Mann–Whitney test. The differences between the efficiency levels in each pair of the other treatments are not statistically significant.

However, as the payoff functions for the forecasting and optimizing tasks were different, it is difficult to draw conclusions from the reported efficiency ratios across some of the treatments. One way to make the results more comparable is to

Table 4
Efficiency for the forecasting and optimizing tasks.

Treatment	Efficiency forecasting (%)	Efficiency optimization (%)	Efficiency index (%)
Treatment 1	87.12	99.85	93.49
Treatment 2	58.99	98.69	78.84
Treatment 3	72.00	95.14	83.57
Treatment 4	77.90	98.68	88.29
Treatment 5	68.75	99.31	84.03

examine implicit production decisions in Treatment 1 (1/6 of the price forecast) and implicit price forecasts in Treatment 2 (6 times the quantity decision), and then calculate the efficiency level of the implicit production decisions in Treatment 1, or the efficiency level of the implicit price forecasts in Treatment 2. Given these numbers we can calculate the efficiency level for both the forecasting and optimizing tasks for all five treatments in a consistent manner and we can define an efficiency index for all the treatments as the mean of the efficiency levels for the two tasks. This index, which allows for efficiency comparisons across the five treatments, is reported in Table 4.

Table 4 reveals that the efficiency level for the implicit optimizing task in Treatment 1 is as high as the comparable efficiency level of the optimizing task in Treatment 2, and sometimes exceeds 100% in the second 25 rounds of the experiment. This suggests that the higher efficiency level reported for Treatment 2 as compared with Treatment 1 may be an artifact of the payoff function differences. Subjects performing the optimization task benefit from small, positive random shocks which result in a higher market price. By contrast, both positive and negative shocks are equally penalizing for subjects performing the forecasting task as both types of shocks lead to higher prediction errors.

Table 4 also reveals that the ranking of the overall efficiency index is now: Treatment 1 > Treatment 4 > Treatment 5 > Treatment 3 > Treatment 2. Decomposing this ranking, for the forecasting task alone the ranking as the same as the overall ranking, while for the optimizing task alone the ranking is: Treatment 1 > Treatment 5 > Treatment 2 > Treatment 4 > Treatment 3. We conducted pairwise Wilcoxon Mann–Whitney tests for differences in market level efficiency for the two tasks and on the efficiency index for periods 1–50. The results indicate that the efficiency level is significantly greater in Treatment 1 for both tasks as well as for the efficiency index as compared to each of all other treatments. The efficiency for the forecasting task is significantly lower in Treatment 2 as compared with the other treatments,[5] but there are no other significant differences in all pairwise comparisons between treatments. As we will see later in the paper, subjects in Treatments 3 and 4 (especially Treatment 3) do not make perfect production decisions given their forecasts. This result suggests that the level of bounded rationality in optimization tasks is not very different in these two treatments. However, if subjects are not fully rational with regard to the optimization task, this may result in inaccurate forecasts resulting in larger forecast efficiency losses.

These results also suggests that high efficiency levels in learning-to-optimize experiments should be treated with some caution; even if efficiency metrics indicate that subjects are doing well on the optimization task, the implicit price forecasts may be far from rational. In this case, the team design with specialized roles provides a clearer view of the efficiency of the decision process for each task (and may improve efficiency levels for forecasting tasks).

4.4. Individual forecasts

The upper panel of Fig. 5 shows the time series of the average individual price forecasts in Treatments 1, 3, 4 and 5 against the REE price. We observe that Treatment 1 (diamonds) converges the fastest followed by Treatment 5 (Xs), Treatment 4 (triangles), and that Treatment 3 (squares) is the slowest to converge (the differences between treatments is significant at the 5% level according to a Wilcoxon Mann–Whitney test). The lower panel of Fig. 5 shows the average variance of individual forecasts over time in Treatments 1, 3, 4 and 5. There, we observe that heterogeneity of forecasts is the largest in Treatment 5 (the difference between Treatment 5 and each of Treatments 1, 3 and 4 is significant at the 5% level), followed by Treatment 3 (the difference between Treatment 3 and each of Treatments 1 and 4 is significant at the 5% level), and there is not much difference between Treatments 1 and 4 (there is no significant difference according to the test). Treatment 5 has the largest variance mainly because of subjects who experimented with very large/small price forecasts in this treatment.

Prior experimental work (Heemeijer et al., 2009) suggests that subjects tend to use simple heuristics in learning to forecast experiments. Two natural candidates that are often used in negative feedback markets (such as the one studied here) are adaptive expectations:

$$p^e_{i,t+1} = p^e_{i,t} + \lambda(p_t - p^e_{i,t}), \tag{8}$$

[5] This result may be due to our assumption that the implicit forecast is 6 times the quantity, or the fact that the subjects do not make conditionally optimal production decisions given their forecast (produce exactly one sixth of the implicit forecast).

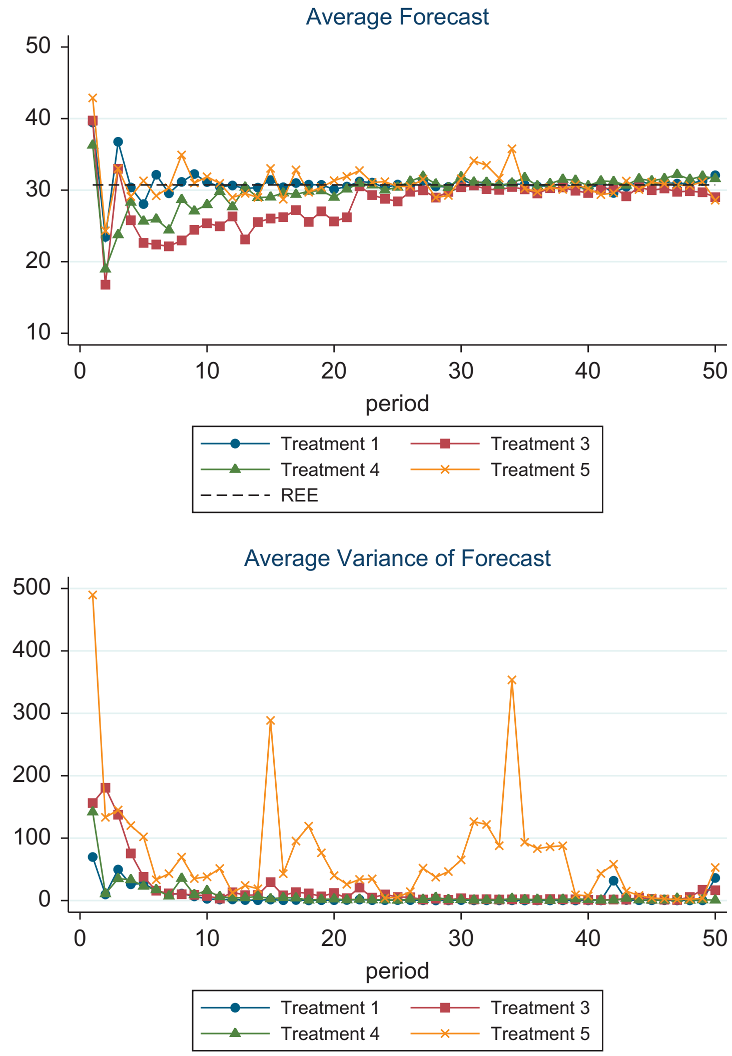

Fig. 5. Average individual forecasts over time in Treatments 1, 3, 4 and 5 (top). Average variance of individual forecasts over time in Treatments 1, 3, 4 and 5 (bottom).

and trend extrapolation rules:

$$p^e_{i,t+1} = p_t + \gamma(p_t - p_{t-1}). \tag{9}$$

The estimated value for γ is typically negative in the negative-feedback market setting that we consider, so we will refer to the trend extrapolation rule (9) as the "contrarian rule" to differentiate this rule from the trend-following version of the same rule where γ is positive. We estimate these two types of rules for each individual subject in our experiment. We call an estimation successful if it generates coefficient estimates that are statistically significant at the 5% level, and if there is no serial correlation in the errors. It turns out that more than 75% of subjects can be successfully characterized by *both* adaptive rules. In those cases we compare the R^2 value for each estimated model and characterize the individual as following the estimated rule having the larger R^2 value. The distribution of individual subjects over the types of forecasting rules is shown in Table 5, while the Tables in Appendix C show the estimation results for the subjects who can be successfully identified using a single rule.

In all three treatments 50% or more subjects can be categorized by the adaptive rule. However, we note that there are relatively more subjects using the adaptive rule in Treatments 1 and 5 than in Treatments 3 or 4. Relating this result to the observed stability of the markets, it would appear that the market price is more stable when there are more subjects using the adaptive rule.

4.5. *Individual supply decisions*

4.5.1. *Descriptive statistics*

The time series of the average supplies in Treatments 2, 3 and 4 are plotted against the REE supply in the top panel of Fig. 6. As with prices, we see that quantity in Treatment 3 (squares) converges toward the REE level in a rather sluggish

Table 5
The fraction of subjects who are characterized by one of the two price forecasting rules (Adaptive or Contrarian) or neither of the two rules.

Treatment	Adaptive (%)	Contrarian (%)	Neither (%)
Treatment 1	66.67	12.50	20.83
Treatment 3	52.38	23.81	23.81
Treatment 4	50.00	27.78	22.22
Treatment 5	83.33	11.11	5.56

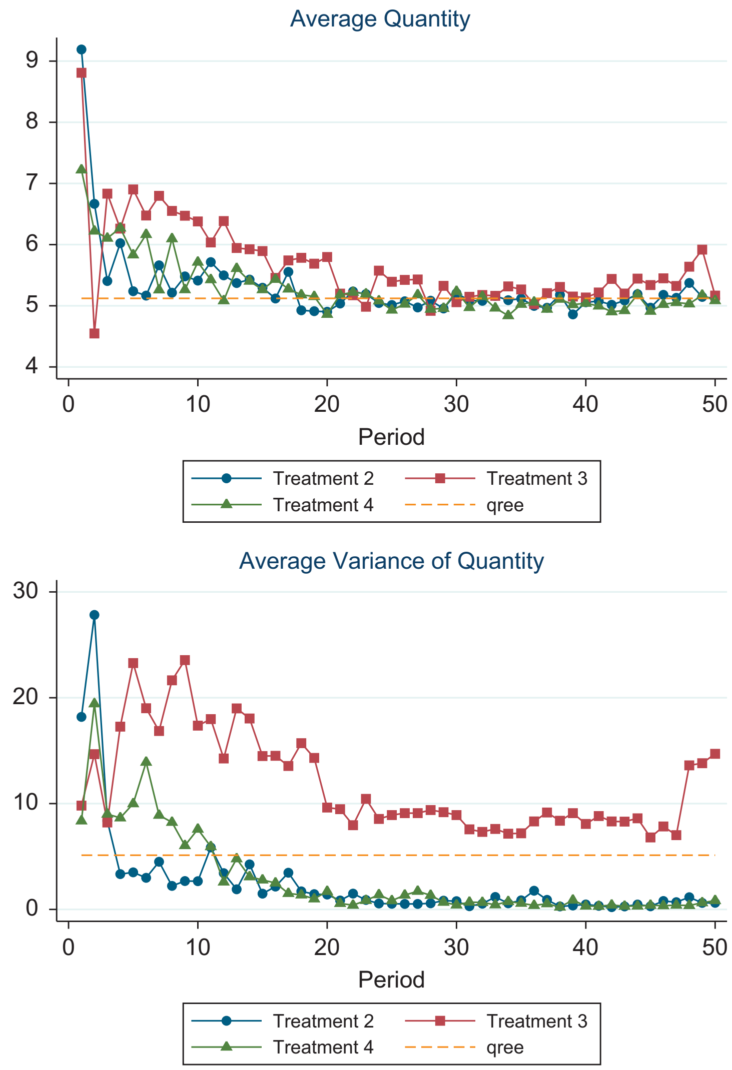

Fig. 6. Average individual supply over time in Treatments 1, 3 and 4 (top). Average variance of individual supply over time in Treatments 1, 3 and 4 (bottom).

manner, and there is not much difference in the average quantity supplied over time between Treatments 2 (diamonds) and 4 (triangles). The bottom panel of Fig. 6 shows the average variance of supply in each treatment. We again observe that the heterogeneity of supply decisions is greatest in Treatment 3, and there is not much difference between Treatments 2 and 4. The difference between Treatment 3 and the other treatments is significant at the 5% level according to a Wilcoxon Mann–Whitney test, and there is no significant difference between Treatments 2 and 4 according to the same test.

4.5.2. Conditional optimality of the supply decision

Recall that if the production manager acts optimally with respect to his own forecast or his partner's forecast, the manager should choose to supply a quantity that is equal to 1/6 of the firm's price prediction, i.e., $S^*(p^e_{h,t}) = p^e_{h,t}/6$. Do

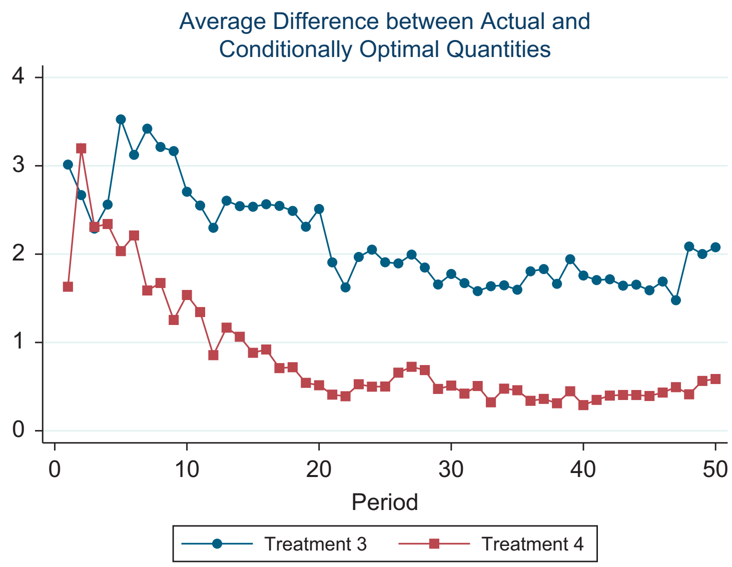

Fig. 7. Average distance between actual supply and the conditionally optimal supply over time in Treatments 3 and 4.

production managers make decisions in this manner? Fig. 7 shows the average difference between the supply chosen by the production manager and the optimal supply given his own or his paired forecaster's forecast in Treatments 3 and 4, respectively. If production managers make decisions optimally, this difference should be zero.

Fig. 7 reveals that production managers in Treatment 4 on average make supply decisions that are closer to the conditionally optimal quantity choice given their partners' price forecast. The average distance between the quantity levels chosen by the subjects and the conditional optimal level ($|q_{h,t} - p^e_{h,t}/6|$) is 1.74 in Treatment 3 and 0.58 in Treatment 4. The difference between the average distance in each market in Treatments 3 and 4 is significant at 5% level according to Wilcoxon–Mann–Whitney test using market averages from each treatment. This finding also suggests that the production managers generally trust their partners. Although trust should not be an issue in Treatment 3, where the forecast and supply decisions are made by the *same* individual, we observe that subjects in Treatment 3 generally fail to make production decisions that are optimal given their *own* price forecasts. We suspect that the reason for this difference in Treatment 3 as compared with Treatment 4 is that doing both tasks (as is required in Treatment 3) is indeed very difficult for a single individual, that is, there is a greater cognitive load in Treatment 3 as compared with Treatment 4. Alternatively, the specialization of tasks in Treatment 4 promotes greater consistency of decisions.[6]

4.5.3. *Individual heterogeneity in Treatment 3*

We have observed a larger variance in individual forecasts and quantity decisions in Treatment 3 as compared with the other three treatments. In this section we delve deeper into the source of this variance focusing on the extent and nature of the heterogeneity among subjects in Treatment 3. In particular, we are ask: (1) are there many or just a few subjects who persistently produce an amount that is far away from the conditionally optimal (CO) level? (2) Is there a correlation between subjects' behavior in the forecasting and quantity choice decisions?

We provide answers to these questions in the following two figures. Fig. 8 is a scatter plot of the average forecast and quantity decision for each individual in Treatment 3. If all individuals forecast the REE price and produce the REE quantity on average, then the points would be concentrated near (30.73, 5.12). If a subject produced on average the conditionally optimal quantity, the associated point should lie on the line labeled $q = f/6$ (f=forecast). A subject is persistently producing too much/too little if his/her point in the quantity dimension is above/below 5.12. A subject is persistently forecasting too high/too low if his/her point in the forecast dimension is to the right/left of 30.73. From Fig. 8, we can see that there are both subjects who persistently produce too much (with average production exceeding 10 in some cases) and too little (average production of about 2) in Treatment 3. The subjects who make these extreme decisions are usually also very far from making conditionally optimal quantity decisions. Importantly we observe that while there is some concentration of choices at (30.73, 5.12), deviations from these REE predictions are widespread in the population of subjects. Some of the deviations are correlated within a market (i.e., markets 1 and 2). For other markets (i.e., markets 5 and 7), there is one individual choosing very high production quantities and it seems that the other subjects in those markets choose low price forecasts and quantities in response to the extreme players' decisions.

[6] To get further at this issue, we estimate a simple model of the form:$|q_{i,t} - p^e_{i,t}/6| = c_0 + c_1 t + \eta_{i,t}$, where the first term is the absolute difference between subject i's chosen quantity and the optimal quantity implied by his/her forecast. The coefficient on the constant term, c_0, is a measure of the size of any persistent bias, while the coefficient c_1 captures how quickly the absolute difference decreases over time. However, the difference between either pair of coefficients, c_0 or c_1, is not statistically significant at the 5% level.

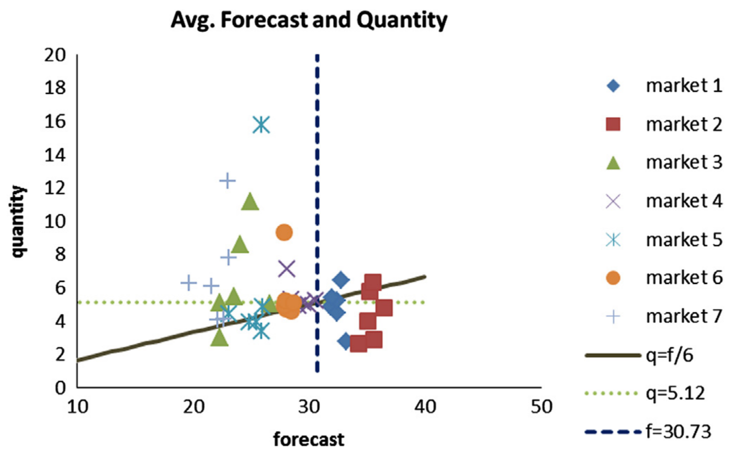

Fig. 8. Scatter plot of the average individual forecast and quantity in Treatment 3.

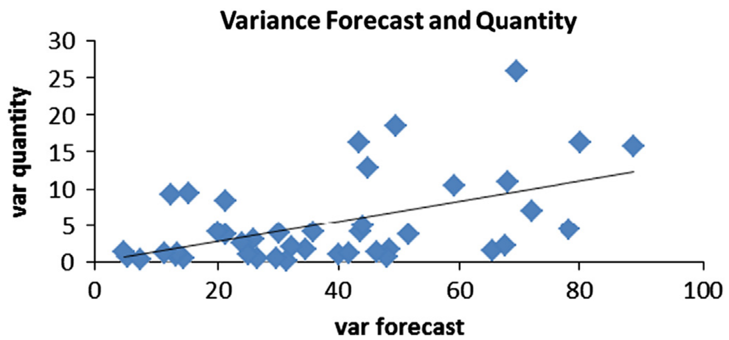

Fig. 9. Scatter plot of the variance of individual forecast and individual supply in Treatment 3.

We next calculated the difference between 1/6 of the average forecast and the average quantity decision made by the same individual. We declare that an individual is making conditionally optimal decisions, on average, if this difference is smaller than 0.5. According to this criterion, we find that only 17 out of 42 subjects (40.5%) are, on average, making conditionally optimal decisions[7]. If we use the same criteria on the data from Treatment 4, we find that 25 of 36 teams/pairs of subjects (69.4%) are on average, making conditionally optimal decisions,[8] which confirms the notion that subjects are more likely to make conditionally optimal decisions in Treatment 4 relative to Treatment 3.

Fig. 9 shows a scatter plot of the variance of individual forecast and quantity decisions in Treatment 3. The figure reveals that there is a positive correlation between the variance in forecasts and the variance in quantity decisions, which suggests that individuals who are less predictable in making forecasts are also more likely to be unpredictable in their quantity decisions. A regression of the variance in quantity choices on the variance in price forecasts, $Variance(q_{i,t}) = c_0 + c_1 Variance(p^e_{i,t}) + \epsilon_i$, yields coefficient estimates $c_0 = 0.1376$, $c_1 = 0.0040$, and the finding that c_1 is significantly different from zero at the 5% significance level.

5. Conclusion

Rational Expectations (RE) macro models have two crucial dimensions: (i) Agents correctly forecast future prices using all available information (i.e., they do not make systematic mistakes) and (ii) Given these expectations, agents solve optimization problems and their solutions (together with market clearing conditions) then determine actual price realizations, that is, there is belief–outcome interaction. These two important dimensions are difficult to simultaneously observe in field data, but have been previously addressed *separately* in learning to forecast experiments (LtFE) and in learning to optimize experiments (LtOE). This paper is the first to explore both dimensions of decision-making in the context of the same expectations-based cobweb model. Specifically, we consider both LtFE and LtOE treatments but we also add two additional treatments where subjects perform both the forecasting or optimizing tasks either independently or as members of a team. Our findings suggest that all four of our main approaches give the same qualitative long-run result,

[7] We denote these subjects using three digit subject numbers, where the first digit is the treatment number, the second one is the market number and the third one is the subject ID on the market. The 17 subjects are 312, 314, 315, 321, 325, 341, 343, 344, 345, 353, 355, 361, 362, 363, 366, 372, 375.

[8] Using the convention described in the previous footnote, these subjects are: 412, 413, 414, 416, 421, 422, 423, 424, 426, 431, 432, 433, 434, 435, 441, 445, 446, 451, 453, 454, 455, 456, 462, 464, 465.

namely convergence to the REE in the context of a cobweb economy with negative feedback, but that the speed of convergence differs across approaches.

Among all the treatments, the LtFE Treatment 1 converges more quickly and reliably than the other treatments. This finding is readily explained by the fact the task in LtFE is easier for subjects than in the LtOE treatment and therefore behavior in LtFE studies should be regarded as an upper bound on the rationality that can be achieved in a laboratory experimental evaluation of RE models. Convergence times under the LtFE design are a little slower if the payoff structure of the LtOE design is used in place of the quadratic loss payoff function that is typically used in LtFE experiments as revealed in our Treatment 5. The combined LtFE+LtOE design of Treatment 3 is the least reliable and slowest to converge to REE. However, Treatment 3 corresponds most closely to what is expected of agents in rational expectations models, i.e., that individual agents are capable of both forecasting *and* optimizing. The current macroeconomic literature on the learning of rational expectations equilibria typically focuses on bounded rationality in agents' expectation formation alone and not on agents' ability to solve optimization problems (see, e.g., Evans and Honkapohja (2001)). The findings reported in this paper suggest that future modeling of the learning of rational expectations equilibria should also consider allowing for bounded rationality in agents' ability to optimize.

At a more microeconomic level, estimation of individual forecast rules suggests that there is not much difference in the price prediction strategies that subjects use across the different treatments of our experiment, although we observe a much higher percentage of adaptive learners in the computer-assisted LtFE Treatment 1. We also find that the use of conditionally optimal supply strategies given price forecasts is more prevalent when subjects interact in teams (Treatment 4) than when they act alone (Treatment 3).

Indeed, we find evidence in support of the notion that “two heads are better than one” in the sense that behavior in Treatment 4 is more rational (closer to RE predictions) than is behavior in Treatment 3, even in the aspect of consistency (how close the production decision is to the conditionally optimal decision for the given price forecast). By contrast with other experimental studies, the two heads in our treatment are specialized in their respective tasks, and this specialization is what may account for the better performance in the team Treatment 4 relative to the individual Treatment 3. Such specialization within a team is also consistent with the real world observation that large financial firms often have separate forecasting and trading departments, and rarely let one department perform the task of the other.

The inconsistency between the price forecast and quantity decision in Treatment 3 can be related to the literature on belief elicitation in game theory experiments. Manski (2002, 2004) argues that eliciting beliefs is always helpful for testing theories in games. Nyarko and Schotter (2003) also find that elicited beliefs provide a better description of game play than inferred beliefs. Rutström and Wilcox (2009), Erev et al. (1993) and Croson (2000) emphasize that eliciting beliefs can change the way people play a game, and Rutström and Wilcox (2009) argue that elicited beliefs may not be a good predictor of individual decisions. The slower convergence in Treatment 3 compared with Treatment 2 suggests that eliciting a forecast may indeed lead subjects to make decisions that depart from the implied optimal level. The low level of conditional optimality in Treatment 3 also suggests that elicited forecasts may be intrusive and alter the associated quantity decision.

The set-up of Treatment 3, where each subject's payment depends on both their price forecast and their quantity choice, is similar to belief-elicitation experiments where subjects receive payoffs from both their stated belief and some belief-conditional action choice. In such environments, risk averse subjects may exhibit “hedging” behavior wherein quantity decisions are made opposite to the direction implied by stated beliefs so as to avoid very low payoffs from large prediction errors. This problem has been analyzed systematically by Blanco et al. (2010), who design an experiment where hedging is possible in one treatment but not in the other. Their main finding is that hedging is not a big concern when hedging opportunities are not very obvious to subjects, such as in a sequential prisoners’ dilemma game, but can be a problem when the hedging opportunities are more transparent, i.e., when subjects play a 2×2 coordination game. Since the subjects in our experiment did not know the number of firms in each market or the price determination equation of the economy, they did not have sufficient knowledge to discover a hedging opportunity. Consequently hedging behavior is unlikely to have played a role in our findings, though it could play a role in future studies using our LtFE+LtOE design if subjects had much more information about the environment in which they were operating. Our Treatment 4, involving specialized forecasting and optimizing roles for teams of players provides one means of potentially disabling hedging opportunities provided, of course, that the members of a team are unable to collude with one another on a hedging scheme.

Our method of comparing and combining LtFE and LtOE experimental methods could also be studied in the context of market environments involving dynamic, state dependent variables, or positive expectation feedback. Positive feedback systems result in greater instability using the LtFE design (e.g., Heemeijer et al., 2009; Bao et al., 2012; Anufriev and Hommes, 2012) due to coordination on trend following strategies. An important question is whether this coordination is also prevalent in LtOE or in combined LtFE+LtOE treatments. We leave such extensions to future research.

Acknowledgments

The authors thank the editor Jörg Oechssler and two anonymous referees for helpful comments. We are grateful to Jasmina Arifovic, Dirk Engelmann, George Evans, Ernst Fehr, Theo Offerman, Andrew Schotter, Lise Vesterlund, Stephanie Wang, seminar participants at the University of Pittsburgh, CESS-CREED Workshop at New York University, the Amsterdam Symposium on Behavioral and Experimental Economics, RES 2012 Annual Conference at the University of Cambridge, 3rd

LICTEM conference at Universitat Pompeu Fabra, ESA 2012 conference at New York University, 18th CEF conference at Prague and 2012 EEA-ESEM meeting for helpful comments and discussions. We acknowledge financial support from the EU 7th framework collaborative project under the grant agreement No. CRISIS-ICT-2011-288501, the Institute for New Economic Thinking (INET) grant for project “Heterogeneous Expectations and Financial Crises” (INO1200026), NSFC project 70932003 and the Research Priority Area in Behavioral Economics of the University of Amsterdam. Te Bao thanks the Willems foundation for funding his visit to the University of Pittsburgh to work on this paper.

Appendices A–D. Supplementary material

Supplementary data and materials associated with this article can be found in the online version at http://dx.doi.org.10.1016/j.euroecorev.2013.04.003.

References

Adam, K., 2007. Experimental evidence on the persistence of output and inflation. Economic Journal 117, 603–636.
Anufriev, M., Hommes, C.H., 2012. Evolutionary selection of individual expectations and aggregate outcomes. American Economic Journal: Micro 4, 35–64.
Arifovic, J., 1996. The behavior of the exchange rate in the genetic algorithm and experimental economies. Journal of Political Economy 104, 510–541.
Bao, T., Hommes, C., Sonnemans, J.H., Tuinstra, J., 2012. Individual expectations, limited rationality and aggregate outcomes. Journal of Economic Dynamics and Control 36, 1101–1120.
Blanco, M., Engelmann, D., Koch, A.K., Normann, H.T., 2010. Belief elicitation in experiments: is there a hedging problem?. Experimental Economics 13 (4), 412–438.
Blinder, A.S., Morgan, J., 2005. Are two heads better than one? Monetary policy by committee. Journal of Money, Credit and Banking 37, 789–811.
Bornstein, G., Yaniv, I., 1998. Individual and group behavior in the ultimatum game: are groups more ‘rational’ players?. Experimental Economics 1, 101–108.
Cason, T., Mui, V., 1997. A laboratory study of group polarization in the team dictator game. Economic Journal 107, 1465–1483.
Cooper, D.J., Kagel, J.H., 2005. Are two heads better than one? Team versus individual play in signaling games. American Economic Review 95, 477–509.
Cox, J.C., 2002. Trust, reciprocity, and other-regarding preferences: groups vs. individuals and males vs. females. In: Zwick, R., Rapoport, A. (Eds.), Advances in Experimental Business Research, Kluwer Academic Publishers, Dordrecht, pp. 331–350.
Crockett, S., Duffy, J., 2012. A dynamic general equilibrium approach to asset pricing experiments. Working Paper.
Croson, R., 2000. Thinking like a game theorist: factors affecting the frequency of equilibrium play. Journal of Economic Behavior and Organization 41 (3), 299–314.
Duffy, J., 2012. Macroeconomics: a survey of laboratory research. Working Paper.
Erev, I., Bornstein, G., Wallsten, T.S., 1993. The negative effect of probability assessments on decision quality. Organizational Behavior and Human Decision Process 55, 78–94.
Evans, G.W., Honkapohja, S., 2001. Learning and Expectations in Macroeconomics. Princeton University Press, Princeton.
Ezekiel, M., 1938. The Cobweb theorem. Quarterly Journal of Economics 52, 255–280.
Heemeijer, P., Hommes, C., Sonnemans, J.H., Tuinstra, J., 2009. Price stability and volatility in markets with positive and negative expectations feedback: an experimental investigation. Journal of Economic Dynamics and Control 33, 1052–1072.
Hommes, C.H., 2011. The heterogeneous expectations hypothesis: some evidence from the lab. Journal of Economic Dynamics and Control 35, 1–24.
Hommes, C.H., Sonnemans, J.H., Tuinstra, J., van de Velden, H., 2005. Coordination of expectations in asset pricing experiments. Review of Financial Studies 18, 955–980.
Hommes, C.H., Sonnemans, J.H., Tuinstra, J., van de Velden, H., 2007. Learning in Cobweb experiments. Macroeconomic Dynamics 11, 8–33.
Huck, S., Normann, H.T., Oechssler, J., 1999. Learning in Cournot oligopoly—an experiment. Economic Journal 109, C80–C95.
Kocher, M., Sutter, M., 2005. When the ‘decision-maker’ matters: individual versus team behavior in experimental ‘beauty contest’ games. Economic Journal 115, 200–223.
Lei, V., Noussair, C.N., Plott, C.R., 2001. Nonspeculative bubbles in experimental asset markets: lack of common knowledge of rationality vs. actual irrationality. Econometrica 69, 831–859.
Lim, S.S., Prescott, E.C., Sunder, S., 1994. Stationary solution to the overlapping generations model of fiat money: experimental evidence. Empirical Economics 19, 255–277.
Lombardelli, C., Proudman, J., Talbot, J., 2005. Committee versus individuals: an experimental analysis of monetary policy decision making. International Journal of Central Banking 1, 181–203.
Manski, Charles F, 2002. Identification of decision rules in experiments on simple games of proposal and response. European Economic Review 46 (4–5), 880–891.
Manski, Charles F, 2004. Measuring expectations. Econometrica 72 (5), 1329–1376.
Marimon, R., Sunder, S., 1993. Indeterminacy of equilibria in a hyperinflationary world: experimental evidence. Econometrica 61, 1073–1107.
Marimon, R., Sunder, S., 1994. Expectations and learning under alternative monetary regimes: an experimental approach. Economic Theory 4, 131–162.
Marimon, R., Spear, S.E., Sunder, S., 1993. Expectationally driven market volatility: an experimental study. Journal of Economic Theory 61, 74–103.
Muth, J.F., 1961. Rational expectations and the theory of price movements. Econometrica 29, 315–335.
Noussair, C.N., Plott, C., Riezman, R., 2007. Production, trade and exchange rates in large experimental economies. European Economic Review 51, 49–76.
Nyarko, Y., Schotter, A., 2003. An experimental study of belief learning using elicited beliefs. Econometrica 70 (3), 971–1005.
Offerman, T., Potters, J., Sonnemans, J.H., 2002. Imitation and belief learning in an oligopoly experiment. Review of Economic Studies 69, 973–997.
Rubinstein, A., 2007. Instinctive and cognitive reasoning: a study of response times. Economic Journal 117, 1243–1259.
Rutström, E.E., Wilcox, N.T., 2009. Stated beliefs versus inferred beliefs: a methodological inquiry and experimental test. Games and Economic Behavior 67, 616–632.
Sargent, T.J., 2008. Evolution and intelligent design. American Economic Review 98, 5–37.
Smith, V.L., Suchanek, G.L., Williams, A.W., 1988. Bubbles, crashes, and endogenous expectations in experimental spot asset markets. Econometrica 56, 1119–1151.
Vega-Redondo, F., 1997. The evolution of Walrasian behavior. Econometrica 65, 375–384.