Hommes, Cars; Lux, Thomas

Working Paper

Individual expectations and aggregate behavior in learning to forecast experiments

Kiel Working Paper, No. 1466

Provided in Cooperation with:
Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: Hommes, Cars; Lux, Thomas (2008) : Individual expectations and aggregate behavior in learning to forecast experiments, Kiel Working Paper, No. 1466, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:
https://hdl.handle.net/10419/24861

Kiel
Working Papers

Kiel Institute
for the World Economy

Individual Expectations and Aggregate Behavior in Learning to Forecast Experiments

by Cars Hommes and Thomas Lux

1466 | November 2008

Web: www.ifw-kiel.de

Kiel Working Paper 1466 | November 2008

Individual Expectations and Aggregate Behavior in Learning to Forecast Experiments

Cars Hommes and Thomas Lux

Abstract:
Models with heterogeneous interacting agents explain macro phenomena through interactions at the micro level. We propose genetic algorithms as a model for individual expectations to explain aggregate market phenomena. The model explains all stylized facts observed in aggregate price fluctuations and individual forecasting behaviour in recent learning to forecast laboratory experiments with human subjects (Hommes et al. 2007), simultaneously and across different treatments.

Keywords: learning, heterogeneous expectations, genetic algorithms, experimental economics.

JEL classification: C91, C92, D83, D84, E3

Cars Hommes
CeNDEF
Department of Quantitative Economics
University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam, The Netherlands
Phone: +31 20-525 4246
E-mail: C.H.Hommes@uva.nl

Thomas Lux
Kiel Institute for the World Economy
24100 Kiel, Germany
Phone: +49 431-8814 278
E-mail: thomas.lux@ifw-kiel.de
E-mail: lux@bwl.uni-kiel.de

Individual Expectations and Aggregate Behavior in Learning to Forecast Experiments*

Cars Hommes

School of Economics *CeNDEF*
University of Amsterdam *E-Mail: C.H.Hommes@uva.nl*

Thomas Lux

Department of Economics *Olshausen Str. 40, 24118 Kiel*
University of Kiel, Germany *E-Mail: lux@bwl.uni-kiel.de*
and
Kiel Institute for the World Economy *Düsternbrooker Weg 120, 24105 Kiel*

November 23, 2008

Abstract: Models with heterogeneous interacting agents explain macro phenomena through interactions at the micro level. We propose genetic algorithms as a model for individual expectations to explain aggregate market phenomena. The model explains all stylized facts observed in aggregate price fluctuations and individual forecasting behaviour in recent learning to forecast laboratory experiments with human subjects (Hommes et al. 2007), simultaneously and across different treatments.

Keywords: learning, heterogeneous expectations, genetic algorithms, experimental economics.

JEL-codes: C91, C92, D83, D84, E3

*Financial support from the EU, through the POLHIA grant no.225408 and Complex Markets grant no.516446 is gratefully acknowledged. We are also particularly grateful to Simone Alfarano, Jasmina Arifovic, Reiner Franke, Juliette Rouchier and Jan Tuinstra for helpful comments and discussions.

1 Introduction

An important feature of models with heterogeneous, interacting agents is that they can explain macro phenomena through simple interactions at the micro level (Kirman, 1993, 2006). Agent-based models have been particularly successful in explaining the main stylized facts of financial markets, such as fat tails and clustered volatility in asset returns (Arthur et al., 1997, Lux and Marchesi, 1999 and Hommes, 2002, among others). Duffy (2006) presents a recent overview of how agent-based models can explain individual behavior and aggregate phenomena in macroeconomics. The main purpose of our paper is to explain aggregate price behavior through interactions of individual learning. In particular, we provide a simple theory of individual learning through genetic algorithms explaining all stylized facts of aggregate price fluctuations in the recent learning to forecast laboratory experiments of Hommes et al. (2007).

Laboratory experiments with human subjects have become an important tool in economic analysis, complementing theoretical, computational and empirical work. A recurring observation from experiments is that individuals often do not behave fully rationally, but tend to use simple heuristics, possibly biased, in making decisions under uncertainty (Tversky and Kahneman, 1974). An extensive bounded rationality research program is developing (e.g. Sargent, 1993) and laboratory experiments are particularly suited to identify behavioral rules that individuals use in economic decision making out of an ocean of potential alternatives (e.g. Kahneman, 2003).

Individual expectations, their interaction and the aggregate outcome they create are at the heart of economics. Duffy (2008), for example, argues that laboratory experiments are important to study the adaptive process by which individuals learn and may or may not enforce convergence to a rational expectations outcome at the macro level. Recently, a number of *learning to forecast* experiments have been conducted to study individual expectation formation and aggregate outcomes, e.g. in Marimon and Sunder (1994), Gerber et al. (2002), Hommes et al. (2005), Sutan and Willinger (2005), Adam (2007) and Heemeijer et al. (2008). In these experiments, subjects must forecast the price of a good, which is determined by market clearing with feedback from individual expectations. Aggregate demand and

supply are computerized, e.g. derived from profit and utility maximization given subjects' individual forecasts. An advantage of this experimental setup is that it provides 'clean data' on expectations, ceteris paribus. These experimental data can therefore be used to test various theories of bounded rationality, individual expectations and learning at the micro level and test how their interaction matches aggregate behavior.

Hommes et al. (2007) conducted learning to forecast experiments in what is perhaps the simplest setting, the classical cobweb 'hog cycle' model describing a standard commodity market with a production lag. Historically, the cobweb model has served as a simple framework to develop and test various expectations hypotheses. Ezekiel (1938) started with naive expectations, Nerlove (1958) advocated adaptive expectations, Muth's seminal paper (Muth, 1961) used the cobweb framework to introduce rational expectations and, more recently, Brock and Hommes (1997), used it to introduce endogenous selection among heterogeneous expectations rules[1].

In their learning to forecast experiments, Hommes et al. (2007) considered three different treatments, a stable, an unstable and a strongly unstable treatment. Stable here refers to the stability of the classical cobweb model under naive expectations[2]. They observed the following three *stylized facts*:

1. the sample mean of realized prices was very close to the RE benchmark *in all three treatments*;
2. the sample variance of realized prices, however, depended on the treatment:
    a) it was close to the theoretical variance of the RE benchmark in the stable treatment, while
    b) it was significantly higher than the RE benchmark (excess volatility) in the unstable and strongly unstable treatments;

[1]The cobweb model has been used in many different applications, ranging from markets for lawyers (Freeman, 1975), engineers (Freeman, 1976), public school teachers (Zarkin, 1985), oil (Krugman, 2001), cattle (Rosen et al., 1994) and beef (Chavas, 2000).

[2]The stability condition states that the ratio between marginal supply and marginal demand at steady state must be smaller than 1 in absolute value (Ezekiel, 1938).

3. in all treatments, realized market prices exhibited no significant linear autocorrelation.

These stylized facts were quite robust over a series of experiments, but they appeared hard to explain by standard learning mechanisms offered by the theoretical literature. While many adaptive learning rules lead to eventual convergence to rational expectations and some other learning rules may generate unstable dynamics and excess volatility, homogeneous expectations models are *unable* to explain the full set of stylized facts *simultaneously* (Hommes, 2009), suggesting that *heterogeneity* of forecasting rules plays a key role in explaining observed aggregate behavior. Hommes (2009) considers some simple examples with two different forecasting rules (e.g. fundamentalists versus naive expectations) and evolutionary competition between them. Although the results in these simple two type heterogeneous expectations examples improve compared to the homogeneous case, they are not capable of explaining *all* stylized facts simultaneously.

In this paper we propose a simple model for micro behavior in order to explain the observed experimental results at the macro level. We model individual learning through genetic algorithms (GAs)[3]. As it turns out, GA experiments with a small population of agents match *all* stylized facts simultaneously across different treatments within a market setting that *exactly* corresponds to that of the laboratory experiments. While it is certainly hard to imagine GAs as an accurate description of human learning in the literal sense, we argue that they may share key properties of the adapta-

[3]Arifovic (1994) used GA learning to explain the cobweb laboratory experiments of Wellford (1989), but there are a number of important differences with our approach. Most importantly, subjects assume the role of producers themselves whereas the subjects of Hommes et al. (2007) have to forecast next period's market price. Besides this major difference in the focus of the experiments, a number of additional differences exist. For example, Wellford used linear demand and supply curves in the experimental setup, implying that the market is either stable or explosive (except for a hairline case). Hommes et al. (2007) use a nonlinear supply curve, so that price dynamics remains bounded while price cycles become a generic possibility. Moreover, Arifovic (1994) only tests for differences of volatility between the stable and unstable treatments, while we match price volatility under GA learning directly to the experiments. Furthermore, Arifovic (1994), for example, did not look at the third stylized fact, the absence of linear predictability. Finally, we also study the average degree of heterogeneity in individual forecasting and how it varies over time.

tions of human subjects when exposed to a new situation that they cannot penetrate theoretically.

We also investigate the *degree of heterogeneity* in individual forecasting behaviour. Heterogeneity in forecasting future asset prices is supported by evidence from stock market survey data, e.g. Vissing-Jorgensen (2003) and Shiller (2000), and in inflation expectations survey data in Branch (2004) and Mankiw et al. (2003). Moreover, in these survey data heterogeneity shows substantial variation through time. Consistent with the findings in the laboratory experiments in our GA learning simulations heterogeneity decreases over time, disappears in the stable treatment but heterogeneity persists in the (strongly) unstable treatment.

Using the GAs for individual learning, our paper makes another contribution that goes beyond the limitations of laboratory experiments. Laboratory experiments are costly, because subjects must be paid according to their performance, and typically experimental markets are small because of capacity limitations. After fitting our GA model to individual learning, we can easily investigate price behavior in alternative, more realistic market scenarios through numerical simulations. In particular, we investigate the occurrence of excess volatility when the number of subjects in the market becomes large and/or when the number of rules per individual becomes large. We also investigate how excess volatility depends on a continuum of parameters such as the ratio of marginal supply and demand.

The paper is organized as follows. Section 2 recalls the learning to forecast experiments, while Section 3 recalls some basic facts of GA-learning. Section 4 compares the stylized facts of the GA simulations to the laboratory experiments, while Section 5 presents simulations of GA learning in more realistic market scenarios. Finally, Section 6 concludes.

2 The forecasting experiment

Hommes et al. (2007) report on a set of cobweb experiments with $K = 6$ participants per session. The participants were asked to predict next period's price under very limited information on the structural characteristics of the market. The realized price p_t in the experiments was determined by

the (unknown) market equilibrium between demand and supply:

$$D(p_t) = \frac{1}{K} \sum_{i=1}^{K} S(p_{i,t}^e) \tag{1}$$

with $p_{i,t}^e$ the price forecast of participant i at time t. We normalize the supply side by dividing by the number of firms to facilitate comparison of settings with different K. Supply $S(p_{i,t}^e)$ was determined by the nonlinear schedule

$$S(p_{i,t}^e) = \tanh(\lambda(p_{i,t}^e - 6)) + 1, \tag{2}$$

while demand was formalized via a simple linear schedule:

$$D(p_t) = a - dp_t + \eta_t \tag{3}$$

with η_t a small stochastic shock drawn from a Normal distribution. Both demand and supply can be derived from profit and utility maximization, and are thus consistent with rational behavior. The resulting equilibrium price is obtained as:

$$p_t = \frac{a - \frac{1}{K} \sum_{i=1}^{K} S(p_{i,t}^e)}{d} + \epsilon_t, \tag{4}$$

where $\epsilon_t = \eta_t / d$. Given the parameters a, d and λ the aggregate realized price p_t depends on individual price expectations as well as the realization of the stochastic shocks.

Participants were only informed about the basic principles of the cobweb-type market. They were advisors to producers, whose only job is to accurately forecast the price of the good for 50 subsequent periods. Pay-offs were defined as a quadratic function of squared forecasting errors, truncated at 0:[4]

$$E = \text{Max}\{1300 - 260(p_{i,t}^e - p_t^*)^2,\ 0\}. \tag{5}$$

Participants were informed that the price would be determined by market clearing and that it would have to be within the range $[0, 10]$. Furthermore, they knew that there was (negative) feedback from individual price forecasts

[4] 1300 points corresponded to 0.5 Euro, so that maximum earings were 25 Euro's. Average earnings ranged from 11.5 to 21 Euro (in about 1.5 hours), over the different treatments.

to realized market price in the sense that if their forecast would increase, the supply would increase and consequently the market price would decrease. Subjects however did not know how large these feedback effects would be, as they had no knowledge of underlying market equilibrium equations. One could say that subjects had *qualitative* information about the market, but no quantitative details.

Participants thus solely had to rely on time series observations and their own behavior vis-à-vis their predictions. Due to the nonlinear aggregation of expectations, the superimposed noise and the ignorance of agents concerning the structural form and parameters, conscious coordination to some kind of rational expectations equilibrium would be extremely demanding if not impossible. This setting is close to the informational assumptions of various theoretical models in the literature on learning and bounded rationality (e.g. Sargent, 1993 and Evans and Honkapohja, 2001) so that the experimental subjects' behavior could be contrasted with various learning rules.

Following the classification of cobweb scenarios under naive expectations, Hommes et al. (2007) distinguished between three treatments. While the parameters of the demand function and the parameters of the noise component remained unchanged across all treatments at $a = 2.3$, $d = 0.25$ and $\epsilon_t = \frac{\eta_t}{d} \sim N(0, 0.5)$, the slope parameter of the supply function was varied over a relatively wide range. Treatment 1 had $\lambda = 0.22$ which under naive expectations would guarantee convergence to the homogeneous rational expectations equilibrium (stable case), treatment 2 had a marginally unstable supply parameter $\lambda = 0.5$ (unstable case), while treatment 3 had a strongly unstable supply parameter $\lambda = 2$ (strongly unstable case). Both the unstable and the strongly unstable treatments lead to a 2-cycle under naive expectations. In all treatments in Hommes et al. (2007) the number of subjects was $K = 6$, but van de Velden (2001) also ran the same experiments in the strongly unstable treatment with $K = 12$ subjects.

Under rational expectations, all individuals would predict $p_t^e = p^*$, that is, they would predict the price at which demand and supply intersect. Given that all individuals have rational expectations, realized prices will be given by

$$p_t = p^* + \epsilon_t. \tag{6}$$

Given the limited market information one can not expect that all individuals have rational expectations at the outset, but one can hope that in such a simple, stationary environment individuals would learn to have rational expectations. The *stylized facts* of these cobweb experiments have already been summarized in the introduction. We briefly recall them here: (1) the sample mean of realized prices is close to the RE benchmark p^* *in all three treatments*; (2) the sample variance of realized prices depends on the treatment: it is close to the RE benchmark in the stable treatment, but significantly higher in the unstable and strongly unstable treatments; (3) realized market prices do not exhibit significant linear autocorrelations. Item (3) indicates that even in the unstable and strongly unstable cases, agents did not leave any linear predictability unexploited. Apparently, the interaction of agents' individual forecasting rules washes out linear predictability in aggregate price behavior. While this points to a certain efficiency of their dispersed effects to predict market prices, market prices did fluctuate 'excessively' in the unstable and strongly unstable treatments. In these cases, price fluctuations exceeded those warranted by the exogenous noise component by more than one order of magnitude so that participants' attempts at learning about the market's behavior did apparently intensify price fluctuations. While these results were quite robust over a series of experiments, they appeared hard to explain by standard learning mechanisms offered by the theoretical literature.

Our goal here is to model individual learning via genetic algorithms (GAs), so that the interaction of these rules produces the stylized facts observed in the experiments simultaneously and across treatments.

3 Learning through Genetic Algorithms

Genetic algorithms have been introduced in the seventies as a stochastic learning algorithm (Holland, 1975). In order to solve complex optimization problems with multiple maxima or minima and possible non-continuities this approach mimics evolutionary operations in nature. One typically starts out with a randomly initiated population of candidate solutions. These initial blind trials are typically encoded as *chromosomes* (strings)

using a binary alphabet[5]. After evaluation of the fitness of the members of the initial population (in terms of the objective function), one applies the genetic operations of *reproduction, crossover*, and *mutation*. For certain reasons, economic applications have mostly added the *election* operator (Arifovic, 1996) as an additional step in the loop of genetic operations between successive generations of individuals. In the following we provide details of these operators and their implementation in the present setting:

1. *Reproduction*: in the transition from one generation to the next, the first step consists in sampling copies of strings from the old generation depending on their *fitness*. In conformity with the pay-off function used in the laboratory experiments fitness was defined as a negative quadratic function of forecast errors with truncation at zero:

$$f_i(t) = \text{Max}\{1300 - 260(p^e_{i,t} - p^*_t)^2,\ 0\}. \tag{7}$$

   The most common reproduction operator is reproduction depending on relative fitness, i.e. copies are sampled from the old population with probabilities $f_i / \sum_j f_j$ biasing the population of new agents towards strategies with higher fitness. Other algorithms in the literature are rank-dependent reproduction or tournament selection in which one draws repeatedly n_1 individuals with replacement from the pool and accepts the $n_2 < n_1$ with the highest fitness from the subsample until the new generation is complete.

2. *Crossover*: when the pool of members of a new generation is complete, genetic material is exchanged between them in order to find new (possibly better) candidate solutions by recombination of the old ones. The simplest version is random selection of a pair of parent strings, determining a cut-off value within the string and sweeping part of the genetic material of the parents when creating their offspring. We follow this approach and take the genetic material of each of both offspring from the left (right) hand side of their 'father' and the right (left) hand side of their 'mother'. This operation takes place

[5] One could as well encode the population as real-coded chromosomes. However, this alternative encounters certain technical problems even if the problem at hand is properly defined for real values (Herrera et al., 1998).

with a probability p_{cross}, while with $1 - p_{cross}$ the parent strings are transferred unchanged into the new generation. We note that both more involved crossover schemes as well as versions with more than two parents can be found in the literature as well.

3. *Mutation*: means that any position (bit) within a chromosome might be flipped into another value (from '0' to '1' or vice versa in the binary alphabet). This happens with a probability p_{mut} once the reproduction and crossover operations are finished.

4. *Election*: in most economic applications, the usual range of genetic operators has been extended by the election algorithm. This compares new chromosomes that have emerged from crossover and mutation with their parents and only admits them to the population if their virtual fitness (measured with respect to the environment in which their parents had to compete) is at least as high as their parents' fitness. This operator serves to prevent agents from adopting clearly inferior strategies. *Most* new strategies that emerge in a genetic process are far off the mark and conscious agents would not voluntarily adopt these new strategies if their trial performance ranks them way below the previous ones.

In many applications of genetic algorithms, the qualitative outcome is largely independent of the particular version of an operator that one adopts (cf. Lux and Schornstein, 2005, for a detailed comparison of various set-ups within a learning context). One may even skip one or the other of the operators (e.g. crossover or election) without changing the overall qualitative results. In our simulations like in various previous economic applications, the results appear to be quite robust under variation of GA parameters and implementations of operators. Unfortunately, one has to rely exclusively on simulations since theoretical results for GAs within an interactive context seem to be essentially out of reach. In our setting with artificially intelligent agents, we tried to reproduce the design of the experiments as close as possible. This applies not only to the parametrization of demand and supply functions, and the choice of a fitness function identical to the payoff function in the laboratory experiments but also to the *number* of agents. Hence we report below experiments conducted with $K = 6$ agents using genetic algorithms to evolve forecasting strategies.

Economic applications of GAs as a learning device have mostly applied it in the sense of 'social learning': the number of agents in these papers equals the number of chromosomes and each agent's chromosome type determines her strategic behavior in the market place (e.g. Arifovic, 1996; Dawid, 1999; Arifovic and Gencay, 2000; Lux and Schornstein, 2005). When the genetic operations are applied to this pool of trader-chromosomes, information is effectively shared and incorporated into the entire new generation via the evolutionary dynamics. This design is certainly at odds with the set-up of the experimental market in which subjects are separated from each other and are not allowed to actively exchange information. We, therefore, assumed that each agent in our computer experiment had his own pool of strategies or forecasting rules which undergo their genetic evolution independently from the rules of other agents (cf. LeBaron et al., 1999, for a similar approach).[6] In our experiments reported below, we endowed each agent with $M = 10$ different chromosomes encoding pairs (α_i, β_i) of the first order autoregressive forecasting rule detailed below. The active rule of each agent, i.e. the rule on which her actual forecast was based, was determined by random draws with probabilities equal to the relative fitness obtained in the last round (which is a monotonic function of the proximity of the forecast to the realized price, cf. eq. 7).

Genetic algorithms require a functional specification of the forecasting rule, whose fitness-maximizing parameter values would then be searched for via the evolutionary algorithm.[7] The simplest specification of a rule would be a constant price forecast. A slightly more complex version would use a constant together with a first order autoregressive component:

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

[6] Vriend (2000) discusses differences between social learning and individual learning in agent-based models.

[7] Genetic programs, in contrast, would allow for the evolution of arbitrary functional specifications using a set of basic functional elements (Chen and Wang, 2002). Since we are able to replicate the experimental stylized facts already with the simpler concept of GAs, we abstain from using the more intricate evolutionary dynamics of genetic programs.

This first order autoregressive (AR1) rule seems a natural forecasting scheme as agents could simply implement it via a linear autoregression using the sample average as their estimate of α_i and the first order sample autocorrelation as the estimate of β_i. Moreover, the AR1 forecasting rule (8) has a simple *behavioral* interpretation, with α_i representing an anchor or observed average price level around which the market price fluctuates, and β_i representing the observed persistence or anti-persistence of price fluctuations[8].

As discussed by Hommes (2009) a representative agent model where all agents employ the same fixed rule, e.g. the rule (8), or where all agents adopt the same adaptive learning scheme, e.g. sample autocorrelation or least squares learning, as a uniform learning mechanism for the whole population, can *not* explain all stylized facts of the experiments simultaneously. A homogenous adaptive learning rule either always enforces convergence to RE (i.e. does *not* explain the second observed stylized fact, the excess volatility in the strongly unstable case) or, in cases where the adaptive learning rule leads to excess volatility, it generates anti-persistent price behavior with significantly negative first order autocorrelation, violating the third stylized fact in the laboratory experiments. In our GA-model, we apply the same functional scheme in a *heterogenous agent framework* with genetically evolved sets of parameters α_i, β_i that could differ across individuals. The key question then is whether the *interaction* of individual forecasting rules (8) can explain all stylized facts observed in aggregate price behaviour simultaneously.

In our simulations the two parameters α_i and β_i are encoded in one string of length $l = 40$, the first (last) 20 bits representing the parameters α_i and β_i, respectively (the number of bits is quite arbitrary and only needs to be large enough for a sufficiently fine-grained structure of the resulting real-valued strategies). α_i is restricted to the interval [0,10] just as in the instructions to participants in the laboratory experiments. The interval for β_i is more arbitrary and had been set symmetrically around zero, $\beta_i \, \epsilon \, [-2, 2]$, allowing for quite strong serial correlation or anti-correlation. The transition from the binary coded evolutionary process to the real-valued forecasts requires

[8]In similar cobweb type laboratory experiments Heemeijer et al. (2007) recently estimated individual forecasting rules, and many individuals actually used forecasting rules of the simple form (8).

to compute:

$$\alpha_{i,t} = 10 \sum_{j=1}^{20} a_{i,t}^{j} \frac{2^{j-1}}{2^{20}-1}, \qquad \beta_{i,t} = -2 + 4 \sum_{j=21}^{40} a_{i,t}^{j} \frac{2^{j-21}}{2^{20}-1} \tag{9}$$

with $a_{i,t}^{j} \in \{0,1\}$ the bits at position $j (j = 1, \ldots, 40)$ of chromosome i at time t.

It is well known that short run GA-simulations may be sensitive to the initialization of the GA's. Therefore, in our (short run) simulations we have chosen an initialization in line with individual forecasts in the first and the second period of the experiment, as discussed in more detail below.

4 Experiments with Genetic Algorithms

In this section we report the results from GA simulations. Unless reported otherwise, each of the six agents will be endowed with $M = 10$ chromosomes and the crossover probability $p_{cross} = 0.6$ (but different values yield similar results). Genetic learning would converge to the rational expectations equilibrium if uniformly across the population all $\beta_{i,t}$ would tend to zero and the $\alpha_{i,t}$ would converge to the RE price p^*. Since the experiments run over a limited number of rounds, an appropriate alignment of our GA simulation with the lab settings is required. Note that in order to start the evaluation of the fitness of agents' strategies, we need two realizations of the market price: the first one serves as the anchor value for the AC strategies in eq. (8) and the subsequent realization serves to evaluate the quality of the AC forecast using eq. (9). As a consequence, evolutionary strategies could be evaluated for the first time at $t = 3$. In order to align the GAs to the lab experiments, we therefore, choose for periods 1 and 2 forecasts and prices from the experiments while the GA population of each agent is initialized randomly. In this way, the 'initial conditions' of the GAs are set equal to those of the lab experiments and our artificial agents initially are subject to the same incentives like the human subjects. As it turns out, this alignment typically guarantees greater similarity than, say, a randomized choice of forecasts at $t = 1$ and $t = 2$ (while still qualitative results are pretty much

the same under different initialization schemes). Obtaining closer proximity by accurate alignment should be seen as an encouraging finding: It is worth emphasizing that this is not a fine-tuning of our algorithm but rather an attempt at matching as closely as possible the experimental scenario.

In our simulations, consistent with the laboratory experiments, we find that the market price fluctuates around the RE benchmark with a sample mean very close to the RE benchmark, but that the level of volatility depends strongly on the treatment. In the stable case, the sample variance is close to its rational expectations benchmark, but it increases significantly beyond the RE benchmark if we proceed to the 'unstable' and 'strongly unstable' scenarios. Fig. 1 shows snapshots from longer simulation runs and Table 1 summarizes some key statistics, for all three treatments, averaged over 1,000 simulations of 50 periods each. The sample mean of individual forecasts (Mean(p^e)) has been obtained by averaging the individual forecasts over all subjects (K=6) and all experiments ($J = 6$) for each treatment. The sample variance of individual forecasts (Var(p^e)) has been computed as follows. Let $\hat{p}^j_{t,k}$ be the price forecast for time period t, by subject k, in experiment j, then the mean forecast for period t in experiment j is $\mu^j_t = \frac{1}{K}\sum_k \hat{p}^j_{t,k}$, with $K = 6$ in the experiments. The sample variance of this mean forecast over all rounds ($T = 50$) of experiment j is given by

$$\mathrm{Var}^j(p^e) = \frac{1}{T-1}\sum_t (\mu^j_t - \frac{1}{T}\sum \mu^j_t)^2. \tag{10}$$

The sample average of individual forecasts can then be obtained by averaging over all experiments ($J = 6$) or over all simulations ($J = 1,000$) respectively, within each treatment:[9]

$$\mathrm{Var}(p^e) = \frac{1}{J}\sum_j \mathrm{Var}^j(p^e). \tag{11}$$

[9]There are other ways of defining the sample mean and sample variance of individual forecasts. For example, one can pool the forecasts for each period t over all individuals i and all experiments j, compute the variance of the pooled forecasts and then average over all rounds. Alternatively, one can start out with the variance over the individual forecasts for each experiment j, for each period t and then average over all experiments and all rounds. The results for these alternative ways of averaging of individual forecasts are very similar to those reported below. In particular, independently of the details of the averaging method, the GA-learning simulations match the laboratory experiments quite nicely.

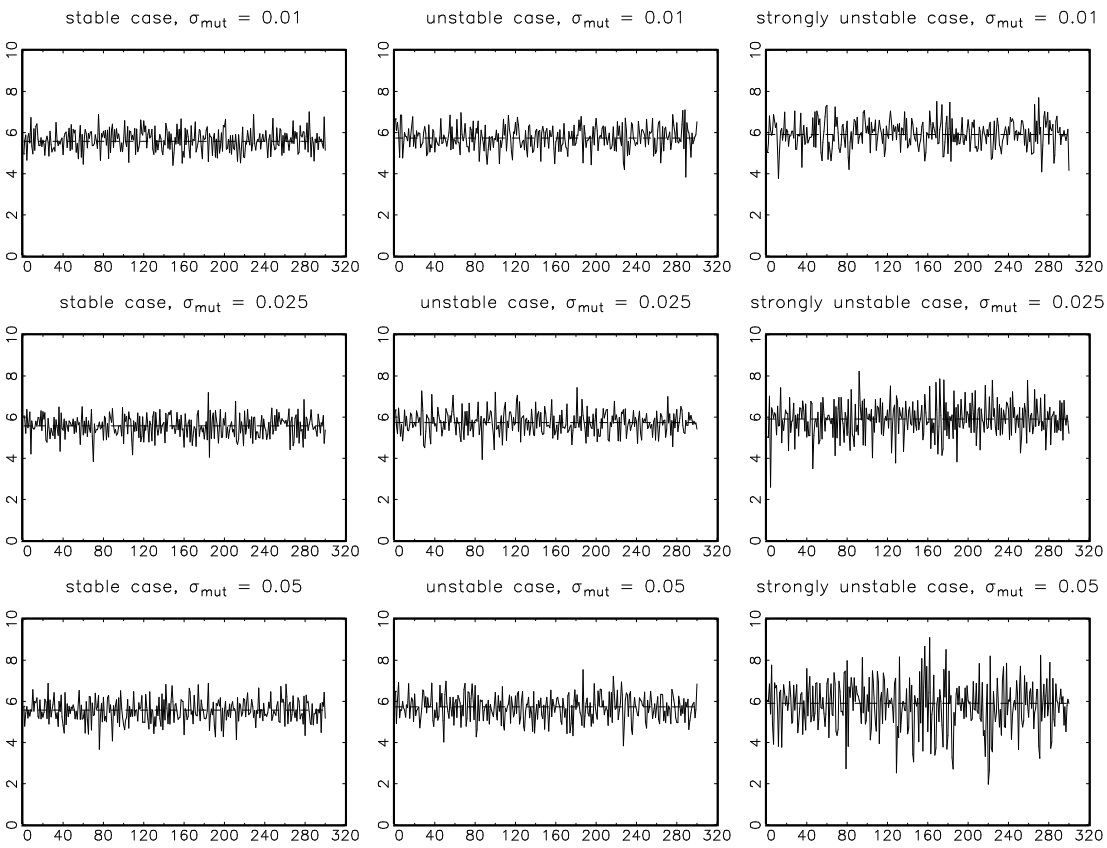

Figure 1: Snapshots from simulations of GA-learning: realized prices (solid) and RE benchmarks (broken lines) for all three treatments (stable, unstable and strongly unstable) and three different values of the mutation probability σ_{mut}.

Table 1: Sample means and sample variances of individual expectations and realized market prices

	Mean(p^e)	Mean(p)	Var(p^e)	Var(p)
Stable treatment				
RE	5.57	5.57	0.042	0.25
Experiments	5.56	5.64	0.087	0.36
$p_{mut} = 0.01$	5.529	5.601	0.083	0.309
$p_{mut} = 0.025$	5.537	5.596	0.095	0.318
$p_{mut} = 0.05$	5.527	5.598	0.115	0.336
Unstable treatment				
RE	5.73	5.73	0.042	0.25
Experiments	5.67	5.85	0.101	0.63
$p_{mut} = 0.01$	5.603	5.841	0.121	0.512
$p_{mut} = 0.025$	5.632	5.808	0.130	0.538
$p_{mut} = 0.05$	5.643	5.795	0.150	0.579
Strongly unstable treatment (K=6)				
RE	5.91	5.91	0.042	0.25
Experiments	5.73	5.93	0.429	2.62
$p_{mut} = 0.01$	5.514	6.154	0.423	1.492
$p_{mut} = 0.025$	5.613	6.026	0.406	1.698
$p_{mut} = 0.05$	5.671	5.923	0.443	2.044
Strongly unstable treatment (K=12)				
RE	5.91	5.91	0.021	0.25
Experiments	5.781	5.937	0.204	1.783
$p_{mut} = 0.01$	5.557	6.155	0.257	0.998
$p_{mut} = 0.025$	5.653	6.037	0.225	1.072
$p_{mut} = 0.05$	5.702	5.946	0.240	1.287

Notes: all parameters have been chosen exactly as in Hommes et al. (2007), i.e., there are $K = 6$ GA agents whose task is to forecast the next period's price. Crossover probability is 0.6 and the election operator is applied. The first and second moments are computed from 1,000 runs of 50 periods each (i.e., using 50,000 observations). The last five rows of the table correspond to the experiments with $K = 12$ in van de Velden (2001) and simulations with $K = 12$ GA agents.

In a similar vein, the variance of realized prices, Var(p) has been computed according to (11), averaging over $J = 1,000$ simulations. Table 1 shows the statistics for the GA simulations, the laboratory experiments and the RE benchmark[10]. The table shows that the GA-simulations are surprisingly close to the laboratory experiments across all treatments. Besides the three treatments of the laboratory experiments, we also distinguish between different settings for the mutation probability, $p_{mut} = 0.01,\ 0.025$ and 0.05, as this appears to be the more interesting aspect of the GA design. As can be seen, price fluctuations also increase *ceteris paribus* with higher mutation probability due to the higher rate of new forecasting rules entering the population. Like in other applications of GAs (cf. Lux and Schornstein, 2005) varying other parameters as well as choosing different specifications of the operators appears to cause no major changes in the overall outcome.

Fig. 2 shows the autocorrelations of prices for the nine scenarios under investigation. All autocorrelations are small, with the first one or two lags exhibiting small negative values, in nice agreement with the laboratory findings (cf. the autocorrelation plots of realized prices in the experiments, Hommes et al., 2007, Fig. 5, pp.21-22). The slight increase of the autocorrelation at the first few lags with higher mutation probability is simply due to the mechanics of a population with high rate of change as the random elements invoke a mean reverting tendency towards the average of a randomized population.

Fig. 3 shows that even under the 'strongly unstable' scenario, the population mean values of the AR1 parameters α_i and β_i are close to their benchmark values under rational expectations: α_i fluctuates in the vicinity of the RE equilibrium price ($p^* = 5.91$) while the average β_i is close to zero. The same applies in the other cases. However, fluctuations around the RE benchmark are stronger for the unstable and strongly unstable cases which leads to stronger fluctuations and excess volatility of market prices consistent with the laboratory experiments.

Overall, the GA experiments with the same parameter setting and incentive structure as in the laboratory experiments appear to closely mimic the set of

[10]For the RE benchmark, Var(p^e) has been computed as the variance of the average of individual forecasts drawn randomly from the RE stochastic process $p^* + \epsilon_t$ in (6), yielding $\mathrm{Var}(p^e) = \sigma^2_{\epsilon_t}/K \approx 0.042$ for $K = 6$.

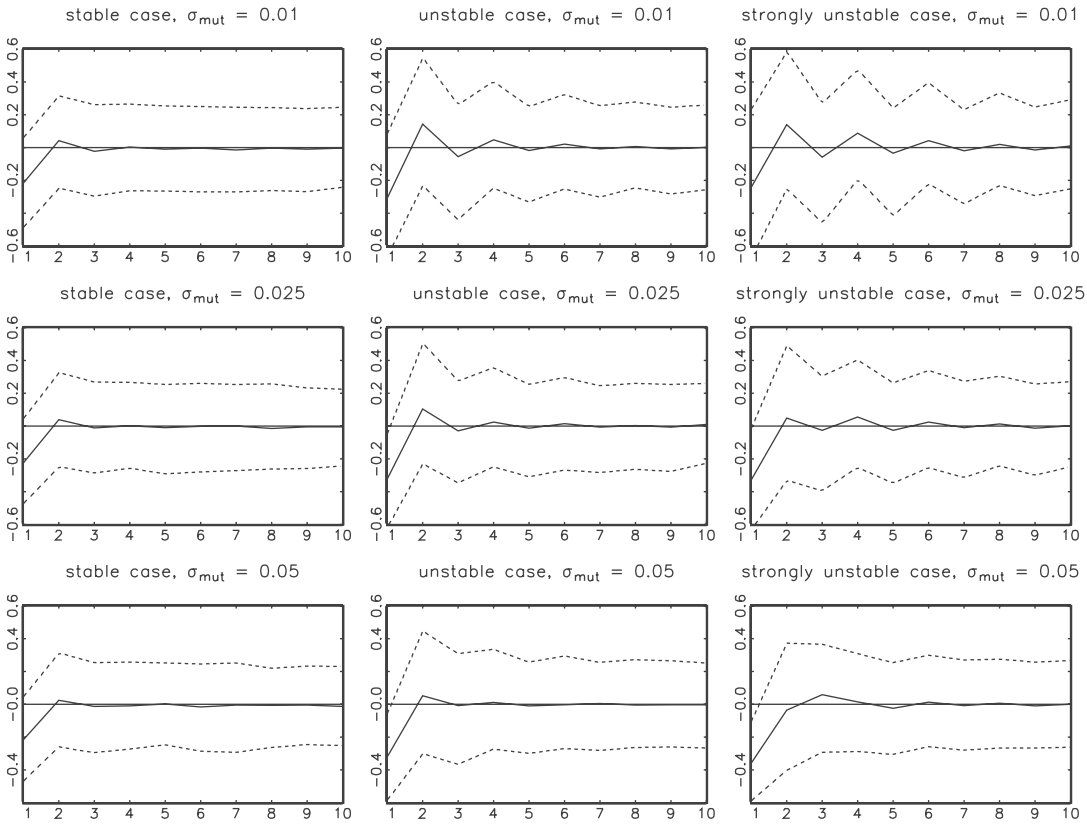

Figure 2: Autocorrelations of realized prices under GA-learning: the process only shows slight negative correlations over one or two lags (consistent with the laboratory experiments). The plots show the means and 95% confidence intervals from 1,000 GA-simulations each extending over 50 periods.

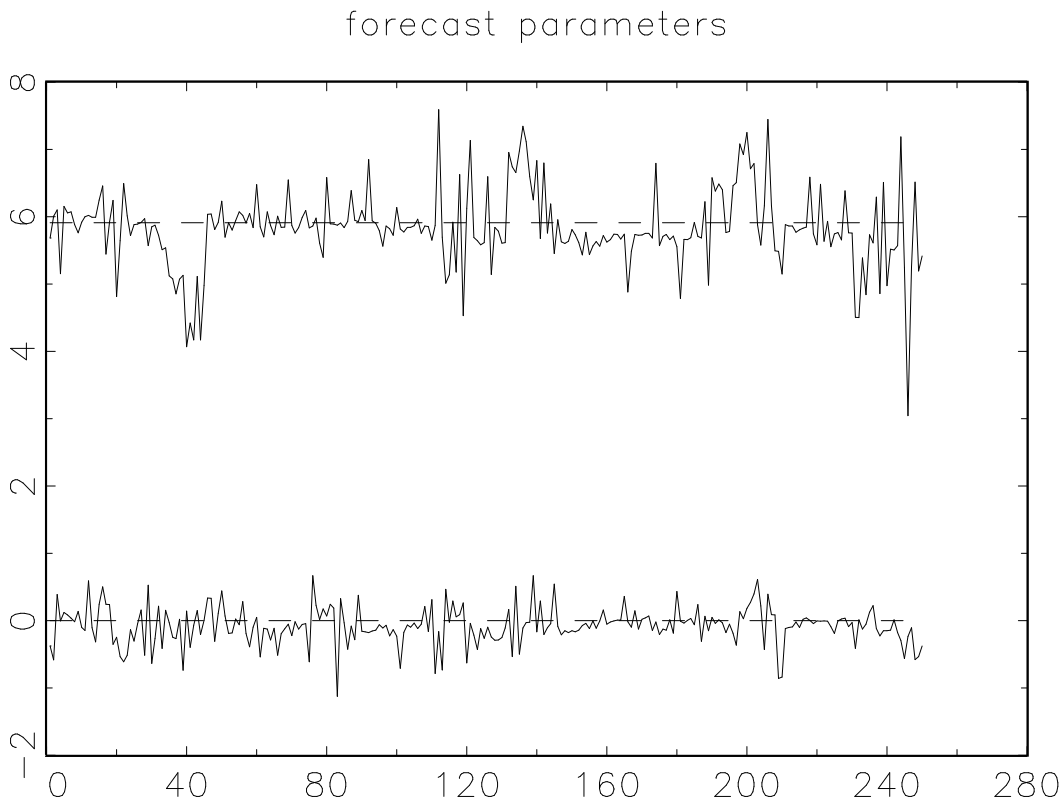

Figure 3: Snapshots of the development of average AR1 parameters α_i and β_i, together with their RE benchmarks $\alpha_i = p^*, \beta_i = 0$ (broken lines), under GA-learning in the 'strongly unstable' treatment.

stylized facts above. We note that the qualitative outcome was quite robust under various modifications of the GA learning mechanism. For example, we get similar results when dispensing with the election operator. The main difference in this setting is that fluctuations in the 'unstable' and 'strongly unstable' treatment become more pronounced and that we see somewhat more significant negative autocorrelation in the first few lags: as mentioned above, this feature can be easily explained by the mean-reverting nature of the dynamics with more random mutations admitted to the population. On economic grounds we might, however, argue that agents should not allow obviously unsuccessful strategies to enter their set of forecast functions (which is why the election operator had been introduced in economic settings) so that we would not place too much weight on these results.

Interestingly, adopting a simpler concept of learning, that dispenses with the AC parameter β_i and restricts forecast rules to a constant α_i also leads to results that share some of the stylized facts. While this scenario leads to similar outcomes for volatility in the three treatments, it is, however, also characterized by anti-persistent price behavior and more significant zigzag patterns of autocorrelations in the strongly unstable case. The simplest GA with only constant rules, therefore, seems to inherit at least part of the oscillating dynamics of the benchmark case of homogenous naive or adaptive expectations. A more intelligent type of forecast rules such as our AR1 rules, taking into account both the average price level and first order autocorrelation, is required to remove linear forecastability. Stated differently, individual learning of the mean alone is not consistent with the laboratory experiments, but some more sophisticated form of individual learning taking into account whether prices are persistent or anti-persistent is needed to remove autocorrelations in aggregate prices. The interaction of individual rules, learning both the price level and the first order autocorrelation, leads to the correct aggregate price level and washes out all autocorrelations in aggregate price fluctuations.

Why do the GA experiments reproduce the experimental results so well? Our conjecture is that GAs and human subjects share the tendency of "learning by experience" and of shifting their strategies towards a specification that would have performed well in the recent past. This is actually the consequence of the built-in genetic operators of GAs. While this leads to

an optimal solution for static problems (at least, if the evolutionary process is allowed to run long enough), with the changing objective functions of an interactive environment it could also lead to repetitive patterns (cf. Lux and Schornstein, 2005). In the absence of structural knowledge about the underlying mechanisms of a decision problem, humans can also still determine what actions or decisions would have performed well in the past. Quite clearly, the laboratory experiments with their unknown forms of the underlying market functions and added stochasticity could not have been fully penetrated by the experimental subjects. However, they could easily focus on the past success and failure of their forecasts and learn to maintain successful strategies. Exploiting the mean together with short-run autocorrelations seems to be one of the simplest strategies that could be pursued in a rule-of-thumb manner without computing the sample autocorrelation exactly (which would normally not be possible given the time pressure of most experimental settings). These rough computations lead to stochastic fluctuations that are similar to the fluctuations caused by the evolutionary dynamics of the GA. The latter feature distinguishes our heterogenous learning scenario from homogenous learning models (Hommes, 2009) which seem unable to explain the full set of stylized facts. Heterogeneity in individual forecasting thus seems to be a key element in explaining all stylized facts at the aggregate level simultaneously.

In summary, our conjecture is that the orientation at successful performance in the past within a reasonable class of forecasting heuristics together with the heterogeneity of the GA design explains its proximity to human behaviour in the lab. We may note that such conformity has also been found in a number of other cases, e.g. in an experimental foreign exchange market (Arifovic, 1996) or public good experiments (Casari, 2004). It is also related to the work of Erev and Roth (1999) on reinforcement learning to explain experiments with repeated games[11]. Anufriev and Hommes (2009) have re-

[11] We also investigated a reinforcement learning (RL) algorithm with similar strategy sets as in our GA setting. Individuals were endowed with AC strategies with parameter space $\alpha_i \in [0, 10]$ and $\beta_i \in [-2, 2]$. Initial parameters were drawn randomly from a uniform distribution and updated with probabilities computed via relative fitness (as in the GA experiments) with fitness defined by the payoff function (5). We distinguished between a myopic RL algorithm, only using the last payoff as fitness function, and a full memory RL scheme that computed fitness as the arithmetic average of all

cently used another form of reinforcement learning to explain learning to forecast experiments in a different, asset pricing framework.

The degree of heterogeneity

In the stable treatment of the laboratory experiments it seems that prices converge to RE and agents learn to coordinate on the RE forecast. In contrast, in the (strongly) unstable treatment prices do not converge, but exhibit excess volatility. Does forecast heterogeneity disappear in the stable treatment and does persistent forecast heterogeneity explain the observed excess volatility in the (strongly) unstable treatment? Whether or not heterogeneity persists is important for economic theory. If beliefs converge to a common rule, long run aggregate price behavior can be described by a representative agent model. If on the other hand beliefs do not converge, heterogeneous expectations models become relevant as a description of short run as well as long run aggregate market behavior. See Hommes (2006) for an extensive discussion of heterogeneous expectations models.

Figure 4 (top panel) shows some typical examples of time series of the six individual forecasts in one group for each of the three treatments. These examples already suggest that heterogeneity quickly disappears in the stable treatment, while heterogeneity is highly persistent in the strongly unstable treatment. Figure 4 (bottom panel) also shows the average degree of heterogeneity, that is, the time development of the standard deviations of individual forecasts ($K = 6$ individuals per group) averaged over all groups in each treatment. More precisely, let $\hat{p}^j_{t,k}$ be the price forecast for time period t, by subject k, in experiment j, then the mean forecast for period t in experiment j is $\mu^j_t = \frac{1}{K}\sum_k \hat{p}^j_{t,k}$. The standard deviation of the mean

previous payoffs. Qualitative results of both settings were not too different, however. The basic outcome of these RL experiments was that (i) agents never converged to the RE benchmark. In particular, the variance was always above the RE benchmark, even in the 'stable' scenario. Mean values were slightly further away from RE prices than under GA learning, (ii) in all cases, realized market prices showed pronounced cyclical patterns indicating that RL agents were not able to exploit all linear structure, (iii) results seemed to be entirely insensitive to the number of strategies (we allowed the strategy set of agents to vary from $M = 50$ over 500 up to 5000). Detailed results are available upon request.

forecast at date t, over $K = 6$ subjects of experiment j is

$$\sigma_t^j = \sqrt{\frac{1}{K-1}\sum_k (\hat{p}_{t,k}^j - \mu_t^j)^2}. \tag{12}$$

The *average degree of heterogeneity* at date t over all experiments ($J = 6$) within a treatment is then defined as the average standard deviation

$$\sigma_t(p^e) = \frac{1}{J}\sum_j \sigma_t^j. \tag{13}$$

The time development of the average degree of heterogeneity in Figure 4 (bottom panel) exhibits two important features: (1) for all treatments forecast heterogeneity decreases over time, and (2) forecast heterogeneity is persistent in the unstable treatment and highly persistent in the strongly unstable treatment.

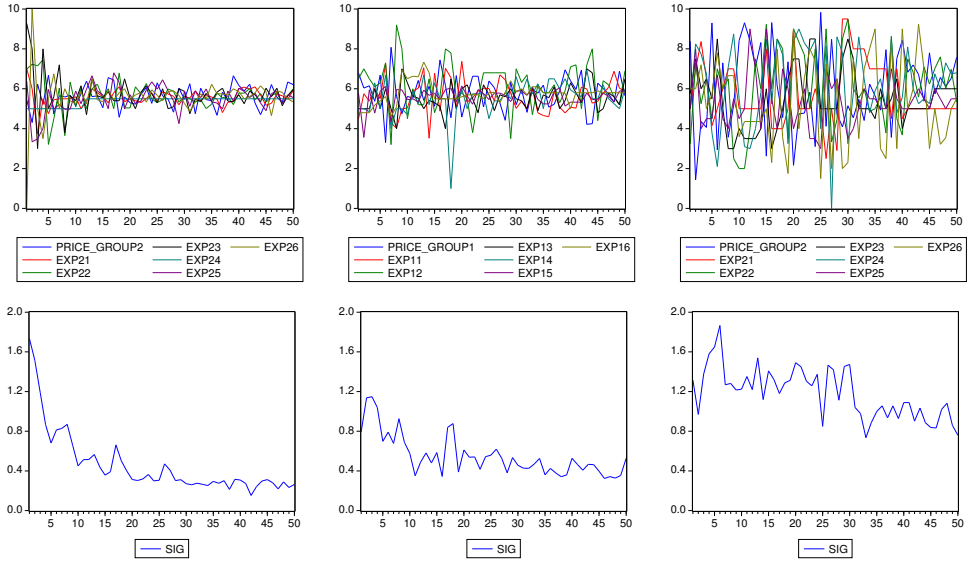

Figure 4: Top panel: Six individual forecasts in one group of the laboratory experiments in the stable (left), unstable (middle) and strongly unstable (right) treatments. Bottom panel: Time development of the average degree of heterogeneity, i.e. the standard deviations of individual forecasts (six individuals) averaged over all (six) groups in the stable treatment (left), the unstable (middle) and the strongly unstable treatment (right).

Figure 5 shows the time development of the average degree of heterogeneity in GA learning simulations in the first 50 periods, averaged over 1000 runs,

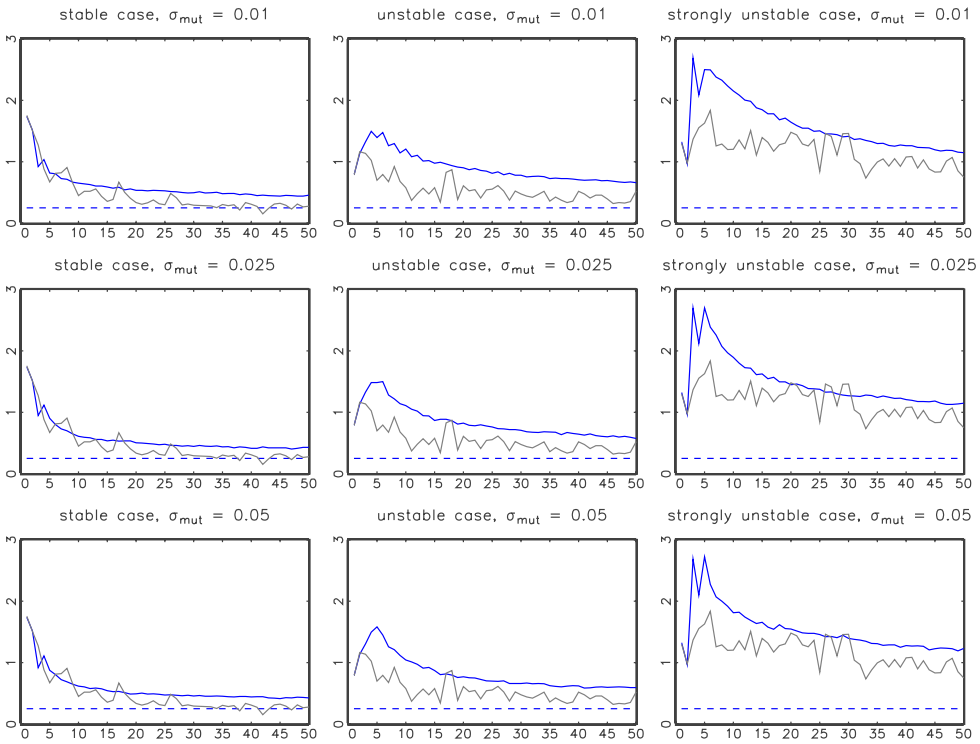

Figure 5: Time development of the average degree of heterogeneity, i.e. the standard deviations of the six individual forecasts in GA learning simulations (dark line) averaged over 1000 runs, in the stable (left), the unstable (middle) and the strongly unstable treatment (right) for different values of the mutation probability p_{mut}. The time series of the average degree of heterogeneity in the corresponding experiments (grey line) as well as the RE benchmark (dotted line) are also shown.

in the stable, the unstable and the strongly unstable treatments. This figure shows that GA-learning simulations reproduce the patterns of the average degree of heterogeneity in the laboratory experiments quite nicely: a quick decrease of forecasting heterogeneity in the stable treatment and a much slower decrease in the (strongly) unstable treatment. In fact, both in the experiments and the GA learning simulations the unstable and strongly unstable treatments exhibit a non-monotonic development of forecasting heterogeneity with an increase in heterogeneity in the early stage of the experiment/simulations due to overshooting and a decrease in heterogeneity after periods $5-7$ due to learning.

5 Beyond the Laboratory Setting

In contrast to experiments with human subjects, additional experiments with genetic algorithms can be conducted at essentially zero cost. In this section we expand our previous experiments into various directions not covered by the laboratory experiments. Among others, we investigate long run price behaviour, how price behaviour depends upon parameter values, in particular the parameter tuning the nonlinearity of the supply curve, and we investigate the consequences of an increase of the number of agents and forecasting strategies in the GA populations.

5.1 Long run behaviour

Table 2 summarizes the long run statistics for all three treatments and three different mutation probabilities, $p_{mut} = 0.01,\ 0.025$ and 0.05. As can be seen, in the stable case the long run average degree of heterogeneity, $\mathrm{Var}(p^e)$, is small and price volatility is quite close to the RE benchmark 0.25. In the unstable treatment price volatility is slightly above the RE variance, while in the strongly unstable treatment the long run price variance is significantly above the RE benchmark due to a larger average degree of heterogeneity. The strongly unstable treatment thus exhibits persistent heterogeneity and long run excess price volatility. Moreover, both the average degree of heterogeneity and excess price volatility increases with the

mutation probability. This seems intuitively clear, since a higher mutation probability leads to a higher rate of new forecasting rules entering the population.

5.2 Parameter sensitivity

In the next set of simulations we explore the transition between the 'nice' price behaviour of the 'stable' treatment and the excessive price volatility of the unstable treatments. Recall that the difference between these treatments is the parameter λ tuning the nonlinearity of the supply curve. We ran the same type of GA experiments with 800 different values of the slope parameters λ ranging from 0.005 to 4 (with increments of 0.005). Fig. 6 reports the mean values and variances of realized prices over 50,000 rounds together with their RE benchmark. It turns out that the variance of realized prices is close to its RE benchmark of 0.25 only for very small values of λ with an almost perfectly linear increase with λ thereafter.[12] In contrast, the average price stays close to its RE benchmark over the whole range of our experiments. While there appears to be a slightly increasing wedge between average price and the RE solution for increasing λ the deviation is always very small compared to the difference between the realization of the second moment and its RE benchmark. We conjecture that this increasing wedge might be more an artifact of our simulation design than a 'true' indication of (small) deviations of the mean price from the rational expectations price. Since p_{RE} is slightly above the center of the admissible range $[0, 10]$ larger fluctuations would generate some asymmetries in realized prices with a slight dominance of lower rather than higher prices. The slight deviation from RE in the first moment (which remains smaller than 2% percent in all scenarios) would then be a numerical consequence of the large deviation in the second moment from its RE benchmark.

[12] A regression of the variance of realized prices on the parameter λ over the second half of our experiments produced a slope parameter 0.12 and a constant 0.25 with R^2 of the regression equal to 0.94. Note that the constant is equal to the variance in RE equilibrium.

Table 2: Long run simulations for different mutation probabilities

	Mean(p^e)	Mean(p)	Var(p^e)	Var(p)
Stable case				
RE	-	5.57	-	0.25
Experiments	5.56	5.64	0.087	0.36
$p_{mut} = 0.01$	5.567	5.575	0.023	0.266
$p_{mut} = 0.025$	5.575	5.574	0.030	0.271
$p_{mut} = 0.05$	5.576	5.565	0.045	0.283
Unstable case				
RE	-	5.73	-	0.25
Experiments	5.67	5.85	0.101	0.63
$p_{mut} = 0.01$	5.723	5.731	0.017	0.292
$p_{mut} = 0.025$	5.729	5.716	0.023	0.313
$p_{mut} = 0.05$	5.725	5.716	0.038	0.355
Strongly unstable case				
RE	-	5.91	-	0.25
Experiments	5.73	5.93	0.429	2.62
$p_{mut} = 0.01$	5.870	5.889	0.019	0.432
$p_{mut} = 0.025$	5.855	5.863	0.054	0.714
$p_{mut} = 0.05$	5.838	5.808	0.151	1.274

Notes: Long run simulations with $K = 6$ GA agents (chromosomes) and different mutation probabilities p_{mut}. The first and second moments for market prices, Mean(p^e) and (Varp^e), are computed from simulations over 50,000 time steps after discarding the first 10,000 observations as transient sample. Mean (p^e) is the mean over the whole simulation of the average forecast across the 6 'agents' in each period. The average degree of heterogeneity, $Var(p^e)$ has been computed according to (10), averaged over $T = 50,000$ periods after a transient of $10,000$ periods.

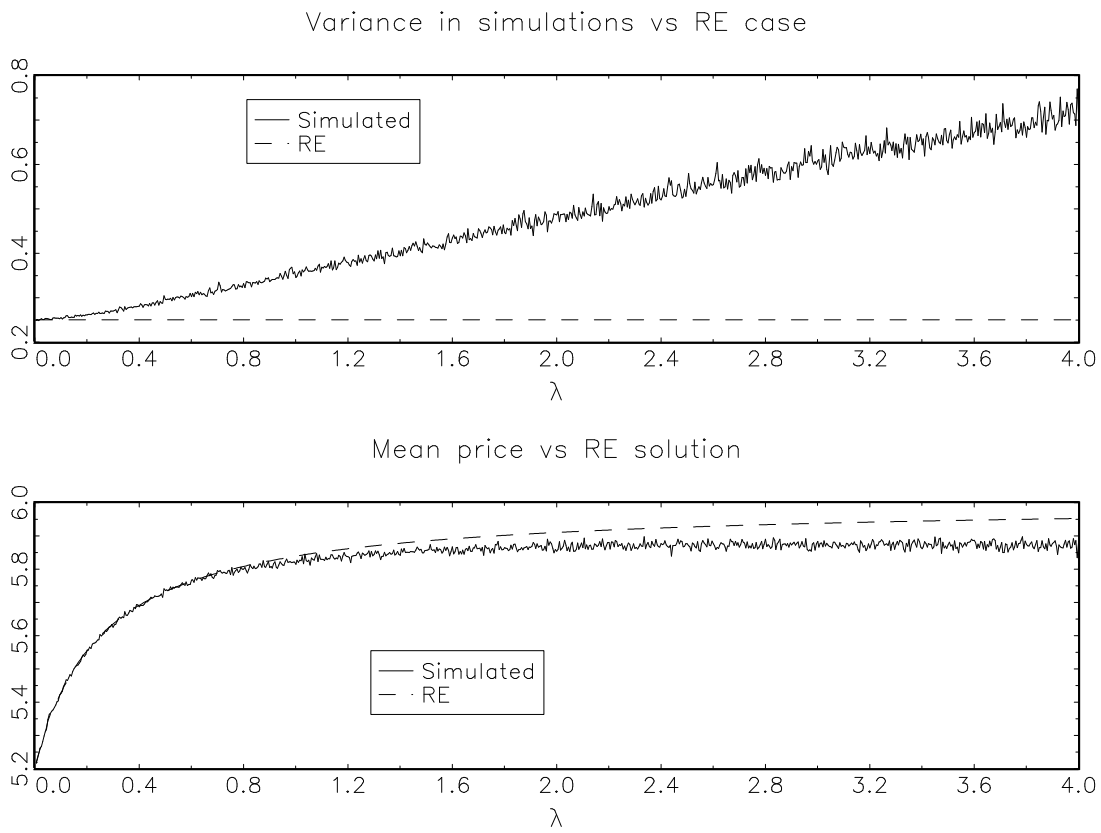

Figure 6: Mean prices (bottom panel) and variances (top panel) from GA simulations of markets with λ ranging from 0.005 to 4. Except for the variation of λ, all parameters are the same as before. Results are sample moments over simulations of 50,000 rounds (after discarding a transient period of 10,000 rounds).

5.3 Simulations with many agents

Another set of Monte Carlo runs investigates what happens if we increase the pool of participants in our forecasting experiments. While laboratory settings are typically restricted to small numbers of agents due to technical restrictions, the availability of subjects and the costs of running large experiments, we can easily extend our previous GA setting to much larger numbers of artificial agents. Since we have normalized supply by dividing through the number of firms in eq. (1), the RE benchmarks for first and second moments remain the same for all population sizes, K.

Table 3 compares the results for population numbers $K \in \{6, 12, 30, 100, 150, 600\}$. Initialization of the GA simulation is done based on the first and second period individual forecasts in the experiments with $K = 6$ subjects in Hommes et al. (2007) as well as the experiments with $K = 12$ subjects in the strongly unstable treatment in van de Velden (2001). For the first period, all experiments (with 6 or 12 subject, with stable, unstable and strongly unstable treatments) have been pooled and the resulting distribution has been fitted with a normal N(5.271, 1.393). Second period forecasts in the experiments differ between treatments, but are very similar for the 6 and 12 subjects cases in the strongly unstable treatment. We, therefore, pooled the forecasts of period 2 over all experiments of each treatment and fitted Normal distributions N(5.279, 1.698) for the stable treatment (only cases with 6 subjects), N(5.952, 1.266) for the unstable treatment (only cases with 6 subjects), and N(6.885, 1.225) for the strongly unstable treatment (pooled over all experiments with 6 or 12 subjects). The above experiments with $K = 6$ to $K = 600$ agents have been initialized by random draws from the pertinent Normal distribution, i.e., the same over all settings in period 1 but the treatment-dependent ones for period 2.

Apparently, higher numbers of agents have a tendency to dampen fluctuations. While there is not much difference in the experiments with stable slope parameter $\lambda = 0.22$, the effect is more pronounced in the unstable and strongly unstable scenarios. The stable case stays close to the RE benchmark for all sizes of the population with the variance of price fluctuations close to the variance of the random term. In the other cases, the excess fluc-

Table 3: Short run simulations for increasing number of subjects

#subjects (K)	Mean(p^e)	Mean(p)	Var(p^e)	Var(p)
Stable case				
$K = 6$	5.531	5.593	0.096	0.316
$K = 12$	5.528	5.597	0.062	0.295
$K = 30$	5.525	5.601	0.042	0.283
$K = 100$	5.527	5.596	0.034	0.275
$K = 150$	5.528	5.599	0.033	0.278
$K = 600$	5.530	5.598	0.031	0.278
Unstable case				
$K = 6$	5.625	5.811	0.140	0.556
$K = 12$	5.643	5.797	0.073	0.417
$K = 30$	5.648	5.793	0.045	0.362
$K = 100$	5.656	5.787	0.034	0.331
$K = 150$	5.652	5.786	0.033	0.326
$K = 600$	5.653	5.788	0.031	0.320
Strongly unstable case				
$K = 6$	5.617	6.023	0.421	1.670
$K = 12$	5.642	6.045	0.264	1.112
$K = 30$	5.659	6.052	0.192	0.821
$K = 100$	5.665	6.051	0.166	0.715
$K = 150$	5.665	6.058	0.168	0.715
$K = 600$	5.666	6.053	0.162	0.694

Notes: Effects of the variation of the number of agents, K. Other parameters are $M = 10$ and $p_{mut} = 0.025$. Increasing K leads to convergence towards the RE benchmark in the stable and unstable treatment, but excess volatility persists in the 'strongly unstable' treatment. The first and second moments are computed from 1,000 runs with 50 periods each (i.e., using 50,000 observation).

Table 4: Long run simulations for increasing number of subjects

#subjects (K)	Mean(p^e)	Mean(p)	Var(p^e)	Var(p)
Stable case				
$K = 6$	5.564	5.575	0.034	0.270
$K = 12$	5.567	5.574	0.020	0.263
$K = 30$	5.570	5.570	0.011	0.259
$K = 100$	5.565	5.571	0.008	0.260
$K = 150$	5.570	5.575	0.008	0.254
$K = 600$	5.566	5.573	0.007	0.255
Unstable case				
$K = 6$	5.722	5.727	0.023	0.314
$K = 12$	5.716	5.734	0.015	0.292
$K = 30$	5.717	5.730	0.009	0.278
$K = 100$	5.715	5.731	0.006	0.265
$K = 150$	5.716	5.729	0.005	0.268
$K = 600$	5.717	5.730	0.005	0.269
Strongly unstable case				
$K = 6$	5.864	5.856	0.052	0.716
$K = 12$	5.882	5.889	0.014	0.441
$K = 30$	5.885	5.889	0.006	0.359
$K = 100$	5.883	5.891	0.004	0.330
$K = 150$	5.884	5.891	0.003	0.320
$K = 600$	5.882	5.890	0.003	0.315

Notes: Long run statistics when the number of agents, K, increases. Other parameters are $M = 10$ and $p_{mut} = 0.025$. Increasing K leads to convergence towards the RE benchmark in all three treatments. Even in the strongly unstable treatment long run market volatility decreases with K, to a value fairly close to the RE benchmark of 0.25. The first and second moments are computed using 50,000 observation, after disregarding the first 10,000 observations as transient.

tuations are clearly reduced when increasing the size of the population. In the 'unstable' case ($\lambda = 0.5$) price volatility decreases from $\text{Var}(p) = 0.556$ for $K = 6$, to $\text{Var}(p) = 0.362$ for $K = 30$, already fairly close to the RE benchmark of 0.25. In the strongly unstable case ($\lambda = 2$) price volatility is reduced by more than 50% for $K = 600$, but is still significantly higher than the RE benchmark ($\text{Var}(p) = 0.694$). In the strongly unstable treatment in the short run, i.e. for the first 50 periods, excess volatility thus persists when increasing the number of agents to $K = 600$.

Another striking feature of these GA-simulations is that an increase of the number of agents beyond $N = 30$ has little effect upon aggregate behavior. In all treatments, price volatility and the average degree of heterogeneity drop significantly when the number of agents is increased from $N = 6$ to $N = 30$, but hardly drop when the number of agents is further increased from $N = 30$ to $N = 600$. This suggests that it may be possible to study macro phenomena in relatively small laboratory experiments with about 30 subjects, a size that is manageable in most experimental laboratories. See also the discussion about the relevance of laboratory experiments in macro in e.g. Duffy (2008).

Table 4 gives an overview of the same statistics in long run simulations, based on 50,000 periods, after a transient of 10,000 periods. Both the stable and the unstable treatments converge to the RE benchmark, with price volatility of $\text{Var}(p) = 0.255$ and $\text{Var}(p) = 0.269$ respectively for $K = 600$. Also the strongly unstable treatment approaches the RE benchmark relatively closely in the long run, although price volatility $\text{Var}(p) = 0.359$ for $K = 30$ and $\text{Var}(p) = 0.315$ for $K = 600$ respectively are still more than 25% above the RE-benchmark.

The laboratory experiments provide a simple and stylized framework that is stationary for 50 periods. In real markets with fluctuating prices, one would perhaps expect larger exogenous shocks to occur occasionally. From this perspective, what we have called the "short run", i.e. the first 50 periods, could be more relevant to real markets than the "long run" where the market is stationary for a very long time.

The decrease of volatility with increasing population is probably easy to explain: adding more agents evokes a law of large numbers. Since our

GA agents are effectively independent stochastic processes, their individual fluctuations should be averaged out when aggregating over more and more individuals.[13] This is what seems to happen in our experiments. Note, in particular the strong decrease of the variance of average price *expectations* in all settings. Of course, if price expectations would converge to the RE benchmark, realized prices would only fluctuate due to the exogenous random noise component.

5.4 Simulations with many rules

Finally, Table 5 reports results of GA-learning simulations where the number of agents has again been fixed to $K = 6$, while the number of rules, M, available to each agent increases from $M = 10$ to $M = 60$.

Note that in a GA setting, the set of rules of an individual need not necessarily be different. In fact, convergence of the GA would imply that the population of rules of an agent becomes fully homogeneous. Increasing M thus does not necessarily mean that an individual has more different rules in each period, but it only increases the potential *sophistication* of the set of rules. As it turns out, at least in the 'strongly unstable' treatment, this higher sophistication leads to an increase of the volatility of realized (as well as expected) prices. With $M = 60$ rules per agent, price volatility has almost doubled from $\mathrm{Var}(p) = 0.824$ to $\mathrm{Var}(p) = 1.546$. In contrast, variation of the number of rules M seems to leave the results of the stable and unstable treatments almost unchanged. We conjecture that the higher number of chromosomes allows agents more easily to react on price fluctuations around the RE benchmark. With a high λ, under naive expectations a small deviation from p_{RE} would lead to a step-wise increase or decrease of the price for some time. Autocorrelation detection by some agents could reinforce this tendency as they would already forestall the direction of the subsequent price changes. With a large number of chromosomes, chances

[13] While there should always be a quasi-deterministic limit for the dynamics of GA populations, this need not necessarily lead to convergence towards some kind of steady state. Lux and Schornstein (2005) provide an example for how adaptive GA agents could converge to perfectly oscillatory dynamics due to their interactions in a large population.

Table 5: Long run simulations for increasing number of rules per agent

# rules (M)	Mean(p^e)	Mean(p)	Var(p^e)	Var(p)
Stable case				
10	5.578	5.565	0.030	0.272
20	5.577	5.567	0.022	0.264
30	5.582	5.569	0.020	0.262
40	5.578	5.567	0.020	0.262
50	5.577	5.570	0.020	0.267
60	5.574	5.570	0.019	0.264
Unstable case				
10	5.729	5.723	0.027	0.328
20	5.731	5.722	0.023	0.315
30	5.734	5.721	0.021	0.310
40	5.734	5.719	0.023	0.308
50	5.737	5.715	0.020	0.312
60	5.735	5.719	0.025	0.316
Strongly unstable case				
10	5.880	5.833	0.092	0.824
20	5.878	5.835	0.156	0.974
30	5.874	5.825	0.226	1.102
40	5.870	5.825	0.308	1.249
50	5.863	5.821	0.418	1.431
60	5.866	5.806	0.468	1.546

Notes: Effects of the variation of the number of rules, M, per GA-agent. Other parameters are $K = 6$ and $p_{mut} = 0.025$. Increasing M seems to leave the results practically unchanged in the 'stable' and 'unstable' cases, but increases both the volatility of predicted prices and realized market prices in the 'strongly unstable' scenario. The moments are extracted from simulations over 50,000 periods after a transient of 10,000 steps.

are increasing to evolve such momentarily advantageous rules. If such rules are admitted to the population, they would enhance fluctuations. It might, therefore, be a mixture of 'naive' adaptation of some agents (modifications of α_i) and trend chasing of others (adapting α_i and β_i) that generates the higher volatility in this case. Unfortunately, a systematic analysis of the interplay between the number of agents and their behavior in experimental settings is beyond the limit of available laboratory resources. Given the autonomous adaptation of human subjects to different environments, it is not clear whether their learning behavior would remain unchanged in groups of different sizes. Our simulations suggest that, at least in the strongly unstable treatment, an increase of the number of learning rules per agent may be a potentially destabilizing force counterbalancing the stabilizing force of an increase in the number of agents in the market. It therefore seems possible, that changes of behavior might compensate for the law-of-large-number tendency in larger groups.

6 Concluding Remarks

Genetic algorithm learning of simple forecasting strategies provides an accurate description of *individual expectations* at the micro level and, at the same time, the *interaction* of these individual rules matches observed aggregate price behavior at the macro level surprisingly well. In the simple framework of the classical cobweb model, the interaction of individual GA-learning rules is able to reproduce all stylized facts in aggregate prices –correct sample mean, excess volatility depending on demand/supply characteristics and no linear predictability – observed in recent learning to forecast laboratory experiments with human subjects. In contrast to homogeneous learning rules, the interaction of heterogeneous GA-learning rules explains all stylized facts simultaneously and across various treatments. It should be emphasized that these results are robust and not sensitive to the GA-specification or the two GA-parameters (the mutation probability p_{mut} and the crossover probability p_{cross}). The GA-algorithms attempt to learn two parameters –the sample mean and the first order autocorrelation coefficient– in a simple AR(1) forecasting rule. Evolutionary selection within a simple class of individual forecasting heuristics, that take into account both the

observed sample mean and the first order sample autocorrelation, thus explains aggregate price behavior surprisingly well.

We have also looked at the average degree of heterogeneity of individual forecasting behaviour. In all treatments, heterogeneity decreases over time. In the stable treatment heterogeneity quickly disappears and the price settles down to its RE benchmark. In the (strongly) unstable treatment heterogeneity decreases somewhat due to learning, but heterogeneity persists, even in the long run. These results suggest that economic theory needs to go beyond representative agent models with homogeneous expectations. The matching of our GA-simulation results with laboratory experiments are consistent with a theory of endogenous selection of heterogeneous expectations, for example, as in Evans and Ramey (1992), Brock and Hommes (1997) and, more recently, in Reis (2006).

Fitting a GA-learning model to the laboratory experiments allows one to go beyond experiments and simulate alternative and more realistic market environments. Through GA-simulations, we have seen that adding more agents to the market has a stabilizing effect, that is, price volatility decreases as the number of agents increases. However, in the strongly unstable treatment, excess price volatility persists when the number of agents becomes large. On the other hand, increasing the potential sophistication by allowing more strategies per individual has a destabilizing effect and makes price behavior more volatile. Additional laboratory experiments could reveal more information about the number of strategies subjects are using, in order to explore which of these two forces will dominate.

We have also seen that an increase in the number of agents beyond 30 has relatively little impact on aggregate price behavior. This suggests that laboratory experiments with 30 interacting subjects may reveal useful information about macro phenomena. Such larger macro experiments as well as applying GA's to other laboratory experiments, in particular other learning to forecast experiments, is a challenge for future work and may shed more light on individual expectation formation, their interaction and aggregate outcomes in alternative market settings.

References

[1] Adam, K., (2007), Experimental evidence on the persistence of output and inflation, *The Economic Journal* 117 (520) , 603–636 .

[2] Anufriev, M. and Hommes, C.H. (2009), Evolution of market heuristics, *Knowledge Engineering Review*, forthcoming.

[3] Arifovic, J., (1994) Genetic algorithm learning and the cobweb model, *Journal of Economic Dynamics and Control* 18, 3-28.

[4] Arifovic, J. (1996), The behavior of the exchange rate in the genetic algorithm and experimental economies, *Journal of Political Economy* 104, 510–541.

[5] Arifovic, J. and R. Gencay (2000), Statistical properties of genetic learning in a model of exchange rate, *Journal of Economic Dynamics and Control* 24, 981–1005.

[6] Arthur, W.B., Holland, J.H., LeBaron, B., Palmer, R. and Tayler, P., (1997) Asset pricing under endogenous expectations in an artificial stock market, in Arthur, W., Lane, D. and Durlauf, S., (eds.) *The economy as an evolving complex system II*, Addison-Wesley, 1997, pp.15-44.

[7] Branch, W.A. (2004), The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations, *Economic Journal* 114, 592-621.

[8] Casari, M. (2004), Can genetic algorithms explain experimental anomalies?, *Computational Economics* 24, 257–275.

[9] Chavas, J.-P. (2000), On information and market dynamics: the case of the U.S. beef market, *Journal of Economic Dynamics and Control* 24, 833-853.

[10] Chen, S.-H. and P. Wang (2002), *Computational Intelligence in Economics and Finance*. Berlin, Springer.

[11] Dawid, H. (1999), *Adaptive Learning by Genetic Algorithms: Analytical Results and Applications to Economic Models*, 2nd ed., Berlin: Springer.

[12] Duffy, J. (2006), Agent-based models and human-subject experiments, in Leigh Tesfatsion and Kenneth L. Judd (ed.), *Handbook of Computational Economics*, edition 1, volume 2, chapter 19, pages 949–1011.

[13] Duffy, J. (2008), Experimental Macroeconomics, Entry in: S. Durlauf and L. Blume, (Eds.), The New Palgrave Dictionary of Economics, 2nd ed., New York: Palgrave Macmillan.

[14] Erev, I. and Roth, A.E., (1999), Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria, *American Economic Review* 88, 848-881.

[15] Ezekiel, M., (1938) The cobweb theorem, *Quarterly Journal of Economics* 52, 255-280.

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

[17] Evans, G.W. and Ramey, G. (1992), Expectation, calculation and macroeconomics dynamics, *American Economic Review* 82, 207-224.

[18] Freeman, R.B. (1975), Legal "cobwebs: a recursive model of the market for new lawyers, *Review of Economics and Statistics* 57, 171-179.

[19] Freeman, R.B. (1976), A cobweb model of the supply and starting salary of new engineers, *Industrial and Labor Relations Review* 29, 236-248.

[20] Heemeijer, P., Hommes, C.H., Sonnemans, J. and Tuinstra, J. (2008), Price Stability and Volatility in Markets with Positive and Negative Expectations Feedback: An Experimental Investigation, *Journal of Economic Dynamics and Control*, forthcoming.

[21] Herrera, F. et al. (1998), Tackling real-coded genetic algorithms: operators and tools for behavioural analysis, *Artificial Intelligence Review* 12, 265–319.

[22] Holland, J. (1975), *Adaptation in Natural and Artificial Systems*, Ann Arbor: University of Michigan Press.

[23] Hommes, C.H., (2002), Modeling the stylized facts in finance through

simple nonlinear adaptive systems, *Proceedings of the National Academy of Sciences* 99, 7221-7228.

[24] Hommes, C.H., 2006. Heterogeneous agent models in economics and finance, In: Tesfatsion, L. and Judd, K.J. (Eds.), Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics, Elsevier, pp.1109-1186.

[25] Hommes, C. (2009), Bounded rationality and learning in complex markets, in: Rosser, B. (ed.), *Handbook of Complexity Research*, in press.

[26] Hommes, C.H., Sonnemans, J., Tuinstra, J., and van de Velden, H., (2005), Coordination of expectations in asset pricing experiments, *Review of Financial Studies* 18, 955-980.

[27] Hommes, C., J. Sonnemans, J. Tuinstra and H. van de Velden (2007) Learning in cobweb experiments, *Macroeconomic Dynamics* 11 (S1), 8-33.

[28] Kahneman, D. (2003), Maps of bounded rationality: Psychology for behavioral economics, *American Economic Review* 93, 1449-1475.

[29] Kirman, A. (1993), Ants, rationality and recruitment, *Quarterly Journal of Economics* 108, 137–156.

[30] Kirman, A. (2006), Heterogeneity in economics, *Journal of Economic Interaction and Coordination* 1, 89-117.

[31] Krugman, P.R. (2001), The oil-hog cycle, *The New York Times*, November 18, p.14.

[32] LeBaron, B., B. Arthur, and R. Palmer (1999), The time series properties of an artifical stock market, *Journal of Economic Dynamics and Control*, 23, 1487–1516.

[33] Lux, T. and Marchesi, M. (1999) Scaling and criticality in a stochastic multi-agent model of a financial market, *Nature* Vol. 397, February 1999, 498–500.

[34] Lux, T. and Marchesi, M. (2000) Volatility clustering in financial mar-

kets: a micro-simulation of interacting agents, *International Journal of Theoretical and Applied Finance* 3, 675-702.

[35] Lux, T. and S. Schornstein (2005), Genetic learning as an explanation of stylized facts of foreign exchange markets, *Journal of Mathematical Economics* 41, 169–196.

[36] Mankiw, N.G., Reis, R.A.M.R. and Wolfers, J. (2003), Disagreement about inflation expectations, *NBER Working Paper* No. W9796.

[37] Marimon, R. and S. Sunder (1994), Expectations and learning under alternative monetary regimes: an experimental approach," *Economic Theory*, 4, 131-162.

[38] Muth, J.F., (1961) Rational expectations and the theory of price movements, *Econometrica* 29, 315-335.

[39] Nerlove, M., (1958), Adaptive expectations and cobweb phenomena, *Quarterly Journal of Economics* 72, 227-40.

[40] Reis, R. (2006), Inattentive producers, *Review of Economic Studies* 73, 793-821.

[41] Rosen, S., Murphy, K. and Scheinkman, J., (1994), Cattle cycles, *Journal of Political Economy* 102, 468-92.

[42] Sargent, T.J., (1993) Bounded rationality in macroeconomics, Clarendon Press, Oxford.

[43] Shiller, R.J., (2000), Measuring bubble expectations and invester confidence, *Journal of Psychology and Financial Markets* 1, 49-60.

[44] Sutan, A. and Willinger, M., (2005), Why do we guess better in negative feedback situations? An experiment of beauty contest games with negative feedback and interior equilibria, Working paper University of Montpellier.

[45] Tversky, A. and Kahneman, D. (1974), Judgment under uncertainty: heuristics and biases, *Science* 185, 1124-1131.

[46] van de Velden, H. (2001), An Experimental Approach to Expectation Formation in Dynamic Economic Systems, Ph-D Dissertation, Tinbergen Institute Research Series 268.

[47] Vissing-Jorgensen, A. (2003), Perspective on behavioral finance: does 'irrationality' disappear with wealth? Evidence from expectations and actions, In: Gertler, M., Rogoff, K. (Eds.) *NBER Macroeconomics Annual*, MIT Press, Cambridge.

[48] Vriend, N.(2000), An illustration of the essential difference between individual and social learning and its consequences for computational analyses, *Journal of Economic Dynamcis & Control*, 24, 1–14.

[49] Wellford, C.P. (1989), A laboratory analysis of price dynamics and expectations in the cobweb model, *Discussion Paper* No. 89-15, Department of Economics, University of Arizona.

[50] Zarkin, G.A. (1985), Occupational choice: an application to the market for public school teachers, *Quarterly Journal of Economics* 100, 409-446.