TI 2013-016/II
Tinbergen Institute Discussion Paper

Individual Expectations and Aggregate Macro Behavior

Tiziana Assenza[1]
Peter Heemeijer[2,3]
Cars Hommes[2]
Domenico Massaro[2]

[1] Catholic University of Milan;
[2] Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute;
[3] ABN AMRO.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Individual Expectations and Aggregate Macro Behavior

Tiziana Assenza[a,b] Peter Heemeijer[c]

Cars Hommes[b,d] Domenico Massaro[b]

January 5, 2013

[a] *ITEMQ, Catholic University of Milan,* [b] *CeNDEF, University of Amsterdam*
[c] *De Nederlandsche Bank,* [d] *Tinbergen Institute, Amsterdam*

First draft: July 15, 2009

Abstract

The way in which individual expectations shape aggregate macroeconomic variables is crucial for the transmission and effectiveness of monetary policy. We study the individual expectations formation process and the interaction with monetary policy, within a standard New Keynesian model, by means of laboratory experiments with human subjects. Three aggregate outcomes are observed: convergence to some equilibrium level, persistent oscillatory behavior and oscillatory convergence. We fit a heterogeneous expectations model with a performance-based evolutionary selection among heterogeneous forecasting heuristics to the experimental data. A simple heterogeneous expectations switching model fits individual learning as well as aggregate macro behavior and outperforms homogeneous expectations benchmarks. Moreover, in accordance to theoretical results in the literature on monetary policy, we find that an interest rate rule that reacts more than point for point to inflation has some stabilizing effects on inflation in our experimental economies, although convergence can be slow in presence of evolutionary learning.

JEL codes: C91, C92, D84, E52.

Keywords: Experiments, Monetary Policy, Expectations, Heterogeneity.

Acknowledgment: We would like to thank conference participants at the Society for Computational Economics, Sydney, July 2009. Financial support from the EU 7^{th} framework collaborative project "*Monetary, Fiscal and Structural Policies with Heterogeneous Agents (POLHIA)*", grant no. 225408, from the Netherlands Organization for Scientific Research (NWO), project "*Understanding Financial Instability through Complex Systems*", and from the EU 7^{th} framework collaborative project "*Complexity Research Initiative for Systemic InstabilitieS (CRISIS)*", grant no. 288501, is gratefully acknowledged. None of the above are responsible for errors in this paper.

1 Introduction

Inflation expectations are crucial in the transmission of monetary policy. The way in which individual expectations are formed, therefore, is key in understanding how a change in the interest rate affects output and the actual inflation rate. Since the seminal papers of Muth (1961) and Lucas (1972) the rational expectations (RE) hypothesis has become the cornerstone of macroeconomic theory, with representative rational agent models dominating mainstream economics. For monetary policy analysis the most popular model is the New Keynesian (NK) framework which assumes, in its basic formulation, a representative rational agent structure (see e.g. Woodford (2003) and Gali (2008)). The standard NK model with a rational representative agent however has lost much of its appeal in the light of empirical evidence: it is clear from the data that this approach is not the most suitable to reproduce stylized facts such as the persistence of fluctuations in real activity and inflation after a shock (see e.g. Chari, Kehoe, and McGrattan (2000) and Nelson (1998)). Economists have therefore proposed a number of extensions to the standard framework by embedding potential sources of endogenous persistence. They have incorporated features such as habit formation or various adjustment costs to account for the inertia in the data (e.g. Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2007)).

In the last two decades adaptive learning has become an interesting alternative to modeling expectations (see e.g. Evans and Honkapohja (1998), Sargent (1999) and Evans and Honkapohja (2001)). Bullard and Mitra (2002), Evans and Honkapohja (2003), Preston (2005) among others, introduce adaptive learning in the NK framework and Milani (2007) shows that learning can represent an important source of persistence in the economy and that some extensions which are typically needed under rational expectations to match the observed inertia become redundant under learning. More recently a number of authors have extended the NK model to include heterogeneous expectations, e.g. Gali and Gertler (1999), Branch and Evans (2006), De Grauwe (2010), Branch and McGough (2009, 2010), Massaro (2012) and Anufriev, Assenza, Hommes, and Massaro (2013).

The empirical literature on expectations in a macro-monetary policy setting can be subdivided in work on survey data and laboratory experiments with human subjects. Mankiw, Reis, and Wolfers (2003) find evidence for heterogeneity in inflation expectations in the Michigan Survey of Consumers and argue that the data are inconsistent with rational or adaptive expectations, but may be consistent with a sticky information model. Branch (2004) estimates a simple switching model with heterogeneous expectations on survey data and provides empirical evidence for dynamic switching that depends on the relative mean squared errors of the predictors. Capistran and Timmermann (2009) show that heterogeneity of inflation expectations of professional forecasters varies over time and depends on the level and the variance of current inflation. Pfajfar and Santoro (2010) measure the degree of heterogeneity in private agents' inflation forecasts by exploring time series of percentiles from the empirical distribution of survey data. They show that heterogeneity in inflation expectations is persistent and identify three different expectations formation mechanisms: static or highly autoregressive rules, nearly rational expectations and adaptive learning with sticky information. Experiments with human subjects in a controlled laboratory environment to study individual expectations have been carried out by, e.g., Marimon, Spear, and Sunder (1993), Marimon and Sunder (1994), Hommes, Sonnemans, Tuinstra, and van de Velden (2005), Adam (2007), Bao, Hommes, Sonnemans, and Tuinstra (2012); see Duffy (2008) for an overview of macro experiments, and Hommes (2011) for an overview of learning to forecast experiments to study expectation formation.

In this paper we use laboratory experiments with human subjects to study the individual expectations formation process within a standard NK setup and fit a theory of heterogeneous expectations to these laboratory data. We ask subjects to forecast the inflation rate under three different scenarios depending on the underlying assumption on output gap expectations, namely fundamental, naive or forecasts from a group of individuals in the laboratory.

In our paper we address the following questions:

- are expectations homogeneous or heterogeneous?
- which forecasting rules do individuals use?
- which monetary policy rules can stabilize aggregate outcomes in learning to forecast experiments?
- which theory of (heterogeneous) expectations and learning fits individual as well as aggregate experimental data?

Our paper makes two contributions. First, we run a learning to forecast experiment within the NK framework and we test the validity of standard monetary policy recommendations (i.e, the Taylor principle) by conducting laboratory experiments with human subjects. The experimental part of our paper is similar in spirit to the learning to forecast experiments of Pfajfar and Zakelj (2010), but differs in at least two crucial ways from their experimental design. While in Pfajfar and Zakelj (2010) participants are forecasting inflation only, we allow, in one of our treatments, agents to forecast both inflation and output gap, in accordance to the theoretical NK model. To our best knowledge, this is the first experimental economy in which fluctuations of the aggregate variables depend endogenously on the individual forecasts of two different variables, inflation and output gap. A second key difference with the experimental design of Pfajfar and Zakelj (2010) concerns the stochastic process of the shocks used in the experiments. In Pfajfar and Zakelj (2010) the shocks follow an AR(1) process, implying an autocorrelated RE solution. In such an environment it is not clear whether fluctuations are expectations driven or driven by economic fundamentals. In contrast, we use small IID shocks to our experimental economy. In the presence of IID shocks, the RE fundamental solution of the model is an IID process, therefore any observed fluctuations in the aggregate variables must be endogenously driven by individual expectations.

The second contribution of our paper is to fit of a *heterogeneous expectations* model to the experimental data. We use the heuristic switching model of Brock and Hommes (1997), and extended by Anufriev and Hommes (2012), to explain the emergence of coordination of individual forecasting rules, and to describe the different aggregate behaviors observed in our New Keynesian experiments.

Another distinguishing feature is that our heterogeneous expectations model is the first model to explain coordination on different forecasting rules for different aggregate variables within the same economy.

The paper is organized as follows. Section 2 describes the underlying NK-model framework, the different treatments, the experimental design and the experimental results. Section 3 proposes a heterogeneous expectations model explaining both individual expectations and aggregate outcomes. Finally, Section 4 concludes.

2 The learning to forecast experiment

In subsection 2.1 we briefly recall the NK model and then we give a description of the treatments in the experiment. Subsection 2.2 gives an overview of the experimental design, while subsection 2.3 summarizes the main results.

2.1 The New Keynesian model

In this section we recall the monetary model with nominal rigidities that will be used in the experiment. We adopt the heterogeneous expectations version of the New Keynesian model developed by Branch and McGough (2009), which is described by the following equations:

$$y_t = \overline{y}^e_{t+1} - \varphi(i_t - \overline{\pi}^e_{t+1}) + g_t \,, \tag{2.1}$$

$$\pi_t = \lambda y_t + \rho\overline{\pi}^e_{t+1} + u_t \,, \tag{2.2}$$

$$i_t = \overline{\pi} + \phi_\pi(\pi_t - \overline{\pi}) \,, \tag{2.3}$$

where y_t and $\overline{y}^e_{t+1}$ are respectively the actual and average expected output gap, i_t is the nominal interest rate, π_t and $\overline{\pi}^e_{t+1}$ are respectively the actual and average expected inflation rates, $\overline{\pi}$ is the inflation target, φ, λ, ρ and ϕ_π are positive coefficients and g_t and u_t are white noise shocks. The coefficient ϕ_π measures the response of the nominal interest rate i_t to deviations of the inflation rate π_t from its target $\overline{\pi}$. Equation (2.1) is the aggregate demand in which the output gap y_t depends on the average expected output gap $\overline{y}^e_{t+1}$ and on the real interest rate $i_t - \overline{\pi}^e_{t+1}$. Equation (2.2) is the New Keynesian Phillips curve according to which the inflation rate depends on the output gap and on average expected inflation. Equation (2.3) is the monetary policy rule implemented by the monetary authority in order to keep inflation at its target value $\overline{\pi}$. The NK model is widely used in monetary policy analysis and allows us to compare our experimental results with those obtained in the theoretical literature. However the NK framework requires agents to forecast *both* inflation and the output gap. Since forecasting two variables at the same time might be a too difficult task for the participants in an experiment we decided to run an experiment using three different treatments. In the first two treatments we make an assumption about output gap expectations (a steady state equilibrium predictor and naive expectations respectively), so that the task of the participants reduces to forecast only one macroeconomic variable, namely inflation. In the third treatment there are two groups of individuals, one group forecasting inflation and the other forecasting output gap. The details of the different treatments are described below.

Treatment 1: steady state predictor for output gap

In the first treatment of the experiment we ask subjects to forecast the inflation rate two periods ahead, given that the expectations on the output gap are fixed at the equilibrium predictor (i.e. $\overline{y}^e_{t+1} = (1-\rho)\overline{\pi}\lambda^{-1}$). Given this setup the NK framework (2.1)-(2.3) specializes to:

$$
\begin{aligned}
y_t &= (1-\rho)\overline{\pi}\lambda^{-1} - \varphi(i_t - \overline{\pi}^e_{t+1}) + g_t\,, && (2.4)\\
\pi_t &= \lambda y_t + \rho\overline{\pi}^e_{t+1} + u_t\,, && (2.5)\\
i_t &= \phi_\pi(\pi_t - \overline{\pi}) + \overline{\pi}\,, && (2.6)
\end{aligned}
$$

where $\overline{\pi}^e_{t+1} = \frac{1}{H}\sum_{i=1}^{H}\pi^e_{i,t+1}$ is the average prediction of the participants in the experiment. Substituting (2.6) into (2.4) leads to the system

$$
\begin{aligned}
y_t &= (1-\rho)\overline{\pi}\lambda^{-1} + \varphi\overline{\pi}(\phi_\pi - 1) - \varphi\phi_\pi\pi_t + \varphi\overline{\pi}^e_{t+1} + g_t\,, && (2.7)\\
\pi_t &= \lambda y_t + \rho\overline{\pi}^e_{t+1} + u_t\,. && (2.8)
\end{aligned}
$$

The above system can be rewritten in terms of inflation and expected inflation:

$$
\pi_t = a + \frac{\lambda\varphi + \rho}{1 + \lambda\varphi\phi_\pi}\overline{\pi}^e_{t+1} + \xi_t\,, \qquad (2.9)
$$

where $a = \dfrac{(1-\rho)\overline{\pi} + \lambda\varphi\overline{\pi}(\phi_\pi - 1)}{1 + \lambda\varphi\phi_\pi}$ is a constant and $\xi_t = \frac{\lambda}{1+\lambda\varphi\phi_\pi}g_t + \frac{1}{1+\lambda\varphi\phi_\pi}u_t$ is a composite shock. Hence, treatment 1 reduces to a learning to forecast experiment on a single variable, inflation, comparable to the learning to forecast experiments on asset prices in Hommes, Sonnemans, Tuinstra, and van de Velden (2005) and on inflation in Adam (2007).[1]

Treatment 2: naive expectations for output gap

In the second treatment we ask subjects to forecast only the inflation rate (two periods ahead), while expectations on the output gap are represented by naive expectations (i.e. $\overline{y}^e_{t+1} = y_{t-1}$). This treatment is similar to the experiment in Pfajfar and Zakelj (2010) who also implicitly assume naive expectations on output.

[1] Given the calibrated values of the structural parameters, described in Section 2.3, the coefficient $\frac{\lambda\varphi+\rho}{1+\lambda\varphi\phi_\pi}$ in (2.9) measuring expectation feedback takes the value of about 0.99 when the policy rule's reaction coefficient $\phi_\pi = 1$, and of about 0.89 when $\phi_\pi = 1.5$. The corresponding expectation feedback coefficient in Hommes, Sonnemans, Tuinstra, and van de Velden (2005) was 0.95.

Given this set up the NK framework (2.1)-(2.3) specializes to:

$$
\begin{aligned}
y_t &= \varphi\overline{\pi}(\phi_\pi - 1) - \varphi\phi_\pi\pi_t + \varphi\overline{\pi}^e_{t+1} + y_{t-1} + g_t\,, &(2.10)\\
\pi_t &= \lambda y_t + \rho\overline{\pi}^e_{t+1} + u_t\,. &(2.11)
\end{aligned}
$$

where $\overline{\pi}^e_{t+1} = \frac{1}{H}\sum_{i=1}^{H}\pi^e_{i,t+1}$ is the average prediction of the participants in the experiment. We can rewrite the above system in matrix form

$$
\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \mathbf{A} + \Omega \begin{bmatrix} 0 & \varphi(1-\phi_\pi\rho) \\ 0 & \lambda\varphi + \rho \end{bmatrix} \begin{bmatrix} \overline{y}^e_{t+1} \\ \overline{\pi}^e_{t+1} \end{bmatrix} + \Omega \begin{bmatrix} 1 & 0 \\ \lambda & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \mathbf{B} \begin{bmatrix} g_t \\ u_t \end{bmatrix} \qquad (2.12)
$$

where $\Omega = (1 + \lambda\varphi\phi_\pi)^{-1}$, $\mathbf{A} = \Omega \begin{bmatrix} \varphi\overline{\pi}(\phi_\pi - 1) \\ \lambda\varphi\overline{\pi}(\phi_\pi - 1) \end{bmatrix}$ and $\mathbf{B} = \Omega \begin{bmatrix} 1 & -\varphi\phi_\pi \\ \lambda & 1 \end{bmatrix}$.

This setup is more complicated than the learning to forecast experiments in Hommes, Sonnemans, Tuinstra, and van de Velden (2005) and Adam (2007) because inflation is not only driven by expected inflation and exogenous noise, but also by the past output gap y_{t-1}. An important difference with Pfajfar and Zakelj (2010) is that we assume IID noise instead of an AR(1) noise process. This assumption helps us to better identify deviations of the experimental outcome from the rational expectations benchmark. In fact, in presence of IID shocks, if fluctuations arise in the experimental inflation process, they must be endogenously driven by expectations. While this logic fully applies for all the others treatments implemented in the experiment, in this specific treatment fluctuations might also arise due to the presence of the backward-looking term on the output gap, y_{t-1}.

Treatment 3: forecasting inflation and output gap

In the third treatment there are two groups of participants acting in the same economy but with different tasks: one group forecasts inflation while the other forecasts the output gap. Agents are divided randomly into two groups, one group is asked to form expectations on the inflation rate and another group provides forecasts on the output gap. The aggregate variables inflation and output gap are thus driven by individual expectations feedbacks from two different variables by two different groups. The model describing the experimental economy can be written as

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \mathbf{A} + \Omega \begin{bmatrix} 1 & \varphi(1-\phi_\pi\rho) \\ \lambda & \lambda\varphi + \rho \end{bmatrix} \begin{bmatrix} \overline{y}^e_{t+1} \\ \overline{\pi}^e_{t+1} \end{bmatrix} + \mathbf{B} \begin{bmatrix} g_t \\ u_t \end{bmatrix}. \tag{2.13}$$

where $\mathbf{A}$, $\mathbf{B}$ and Ω are defined as in treatment 2, while $\overline{y}^e_{t+1} = \frac{1}{H}\sum_{i=1}^{H} y^e_{i,t+1}$ and $\overline{\pi}^e_{t+1} = \frac{1}{H}\sum_{i=1}^{H} \pi^e_{i,t+1}$ are respectively the average output gap and the average inflation predictions of the participants in the experiment. As already pointed out, in treatments 1 and 2 individuals are asked to forecast only the inflation rate two periods ahead, assuming respectively that the expected future output gap is given by the equilibrium predictor ($\overline{y}^e_{t+1} = (1-\rho)\overline{\pi}\lambda^{-1}$) or follows naive expectations ($\overline{y}^e_{t+1} = y_{t-1}$). An important novel aspect of Treatment 3 is that our experimental economy is driven by individual expectations on *two* different aggregate variables that interact within the NK framework.

Treatments a/b: passive versus active monetary policy

In order to study the stabilization properties of a monetary policy rule such as (2.3), we ran two experimental sessions for each of the three different treatments described above. In session "a" the monetary policy responds only weakly to inflation rate fluctuations i.e., the Taylor principle does not hold ($\phi_\pi = 1$), while in session "b" monetary policy responds aggressively to inflation i.e., the Taylor principle holds ($\phi_\pi = 1.5$).[2]

Table 1 summarizes all treatments implemented in the experiments. In total 120 subjects participated in the experiment in 16 experimental economies, 3 for each of the treatments $1a$, $1b$, $2a$, and $2b$ with 6 subjects each, and 2 experimental economies for treatments $3a$ and $3b$ with 12 subjects each. Total average earnings over all subjects were € 32.

[2]Notice that when the policy parameter ϕ_π is equal to 1, the system in Treatments 2 and 3 exhibits a continuum of equilibria.

	ϕ_π	π^e_{t+1}	y^e_{t+1}	# groups	average earnings $\pi\,(y)$ in €
Treatment 1a	1	$\overline{\pi}^e_{t+1}$	$(1-\rho)\overline{\pi}\lambda^{-1}$	3	31
Treatment 1b	1.5	$\overline{\pi}^e_{t+1}$	$(1-\rho)\overline{\pi}\lambda^{-1}$	3	37
Treatment 2a	1	$\overline{\pi}^e_{t+1}$	y_{t-1}	3	28
Treatment 2b	1.5	$\overline{\pi}^e_{t+1}$	y_{t-1}	3	36
Treatment 3a	1	$\overline{\pi}^e_{t+1}$	$\overline{y}^e_{t+1}$	2	28 (28)
Treatment 3b	1.5	$\overline{\pi}^e_{t+1}$	$\overline{y}^e_{t+1}$	2	34 (32)

Table 1: Treatments summary

2.2 Experimental design

The experiment took place in the CREED laboratory at the University of Amsterdam, March-May 2009. For treatments 1 and 2, groups of six (unknown) individuals were formed who had to forecast inflation two periods ahead; for treatment 3 two groups of six individuals were formed, one group forecasting inflation, the other group forecasting the output gap. Most subjects are undergraduate students from Economics, Chemistry and Psychology. At the beginning of the session each subject can read the instructions (see Supplementary material, (Translation of Dutch) Instructions for participants) on the screen, and subjects receive also a written copy. Participants are instructed about their role as forecasters and about the experimental economy. They are assumed to be employed in a private firm of professional forecasters for the key variables of the economy under scrutiny i.e. either the inflation rate or the output gap. Subjects have to forecast either inflation or the output gap for 50 periods. We give them some general information about the variables that describe the economy: the output gap (y_t), the inflation rate (π_t) and the interest rate (i_t). Subjects are also informed about the expectations feedback, that realized inflation and output gap depend on (other) subjects' expectations about inflation and output gap. They also know that inflation and output gap are affected by small random shocks to the economy. Subjects did *not* know the equations of the underlying law of motion of the economy nor did they have any information about its steady states. In short, subjects did not have quantitative details, but only qualitative information about the economy, which is a standard strategy in learning to forecast experiments (see Duffy (2008) and Hommes (2011)).

The payoff function of the subjects describing their score that is later converted into Euros is given by

$$score = \frac{100}{1+f}, \tag{2.14}$$

where f is the absolute value of the forecast error expressed in percentage points. The points earned by the participants depend on how close their predictions are to the realized values of the variable they are forecasting. Information about the payoff function is given graphically as well as in table form to the participants (see Fig. 1). Notice that the prediction score increases sharply when the error decreases to 0, so that subjects have a strong incentive to forecast as accurately as they can; see also Adam (2007) and Pfajfar and Zakelj (2010), who used the same payoff function.

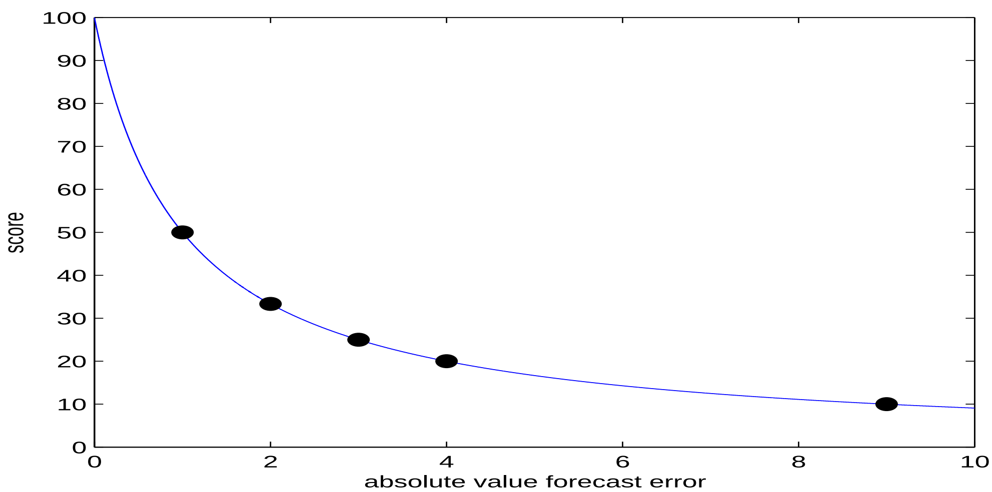


Figure 1: Payoff function

Absolute forecast error	0	1	2	3	4	9
Score	100	50	$\frac{33}{3}$	25	20	10

In each period individuals can observe on the left side of the screen the time series of realized inflation rate, output gap and interest rate as well as the time series of their own forecasts. The same information is displayed on the right hand side of the screen in table form, together with subjects own predictions scores (see Fig. 2). Subjects did not have any information about the forecasts of others.

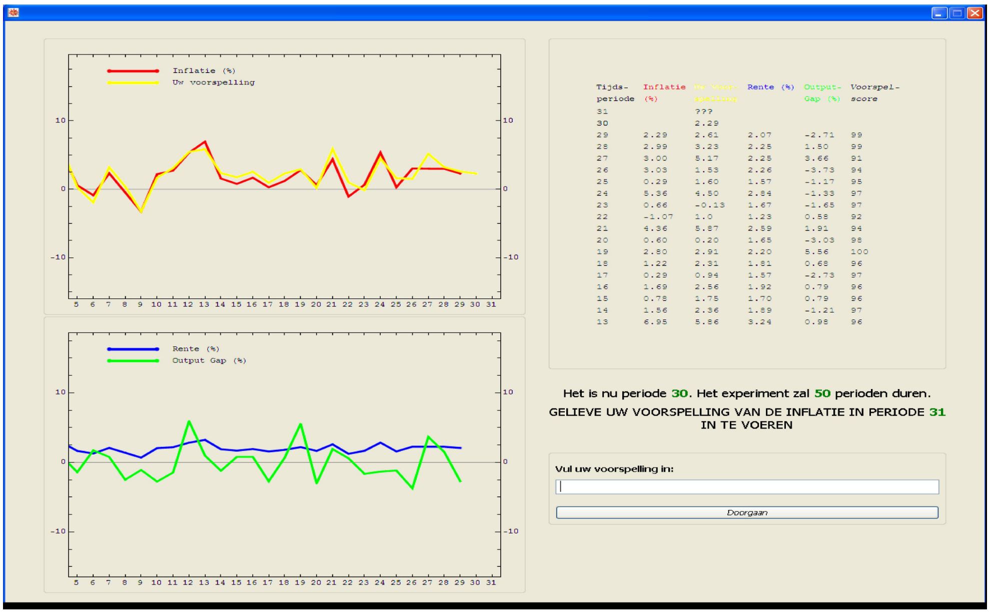


Figure 2: Computer screen for inflation forecasters with time series of inflation forecasts and realizations (top left), output gap and interest rate (bottom left) and table (top right).

2.3 Experimental results

This subsection describes the results of the experiment. We fix the parameters at the Clarida, Gali, and Gertler (2000) calibration, i.e. $\rho = 0.99$, $\varphi = 1$, and $\lambda = 0.3$, and we set the inflation target to $\overline{\pi} = 2$.

Fig. 3 depicts the behavior of the output gap, inflation and individual forecasts in the three different sessions of treatments $1a$ and $1b$ with output expectations given by the steady state predictor. The dotted lines in the figures represent the RE steady states for inflation and output gap that are respectively 2 and 0.07. In treatment $1a$ ($\phi_\pi = 1$) we observe convergence to a non-fundamental steady state for two groups, while the third group displays highly unstable oscillations.[3]

[3]The unstable fluctuations were mainly caused by one participant making very high and very

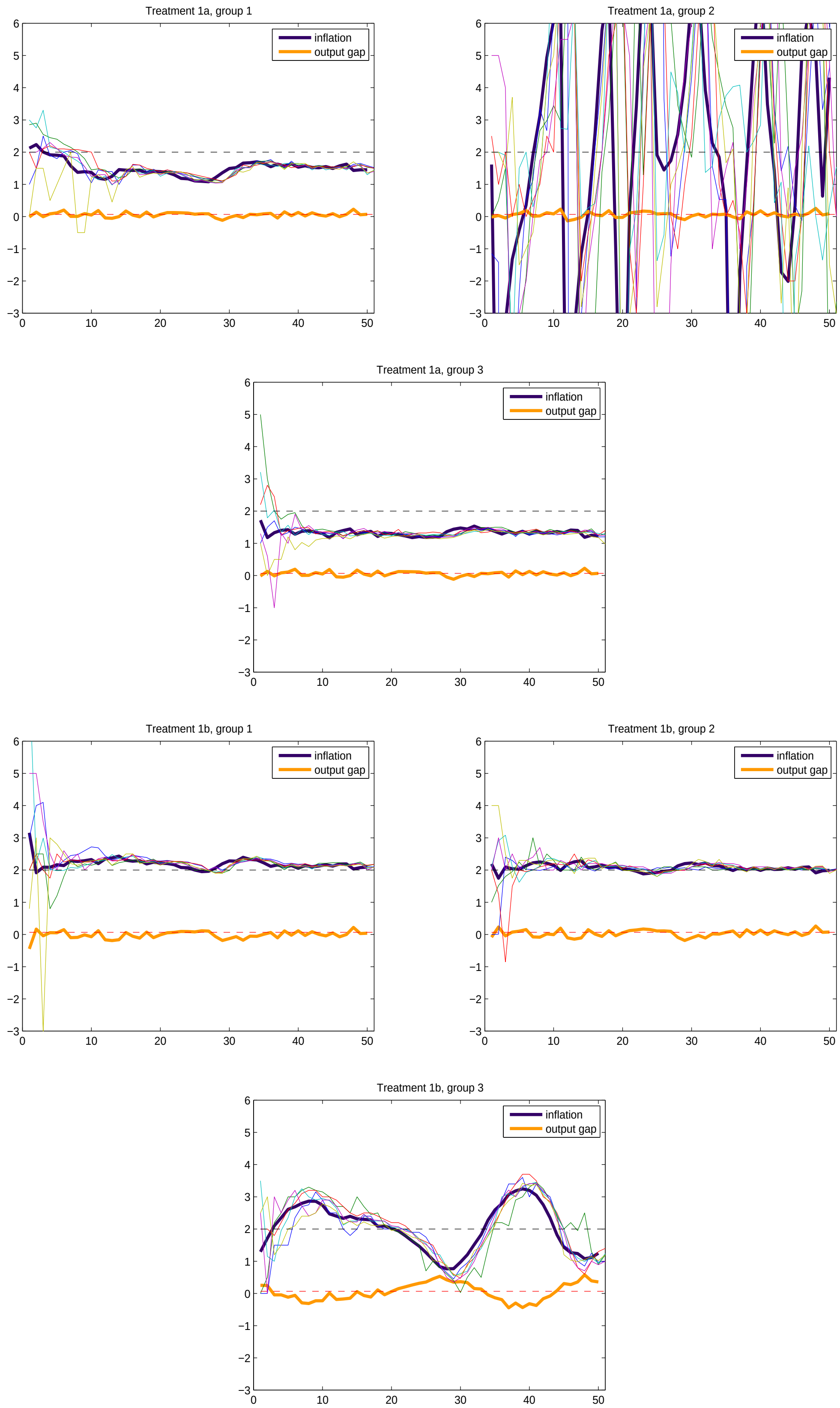


Figure 3: Time series of Treatment 1, with fundamental predictor for the output gap. **Upper panels**: Treatment 1a ($\phi_\pi = 1$). **Lower panels**: Treatment 1b ($\phi_\pi = 1.5$). Blue thick line: realized inflation; yellow thick line: realized output gap; thin lines: individual forecasts for inflation.

In treatment $1b$ ($\phi_\pi = 1.5$) we observe convergence to the inflation target for two groups, while the third group exhibits oscillatory behavior which is by far less pronounced than what we observed in treatment $1a$, group 2.

We conclude that, under the assumption of a fundamental predictor for expected future output gap, a more aggressive monetary policy that satisfies the Taylor principle ($\phi_\pi > 1$) stabilizes inflation fluctuations and leads to convergence to the desired inflation target in two of the three groups.

Fig. 4 shows the behavior of the output gap, inflation and individual forecasts in three different groups of treatments $2a$ and $2b$ with naive output gap expectations. In treatment $2a$ ($\phi_\pi = 1$) we observe different types of aggregate dynamics. Group 1 shows convergence to a non-fundamental steady state. Group 2 shows oscillatory behavior with individual expectations coordinating on the oscillatory pattern. In this session the interest rate hits the zero lower bound in period 43 and the experimental economy experiences a phase of decline in output gap but eventually recovers. In group 3 the behavior is even more unstable: inflation oscillates until, in period 27, the interest rate hits the zero lower bound and the economy enters a severe recession and never recovers. In treatment $2b$ ($\phi_\pi = 1.5$) we observe convergence to the fundamental steady state for two groups, while the third group exhibits small oscillations around the fundamental steady state.

We conclude that also under the assumption of naive expectations for the output gap, an interest rate rule that responds more than point to point to deviations of the inflation rate from the target stabilizes the economy.

The upper panels of Fig. 5 reproduce the behavior of the output gap, inflation and individual forecasts for both variables in two different sessions of treatment $3a$. Recall that in treatment 3 realized inflation and output gap depend on the individual forecasts for both inflation and output gap. In both groups of treatment $3a$ ($\phi_\pi = 1$) we observe (almost) convergence to a non-fundamental steady state.[4] In

low forecasts.

[4]Note that group 1 ends in period 26 because of a crash of one of the computers in the lab. Moreover realized inflation and output gap in group 2 are plotted until period 49 because of an

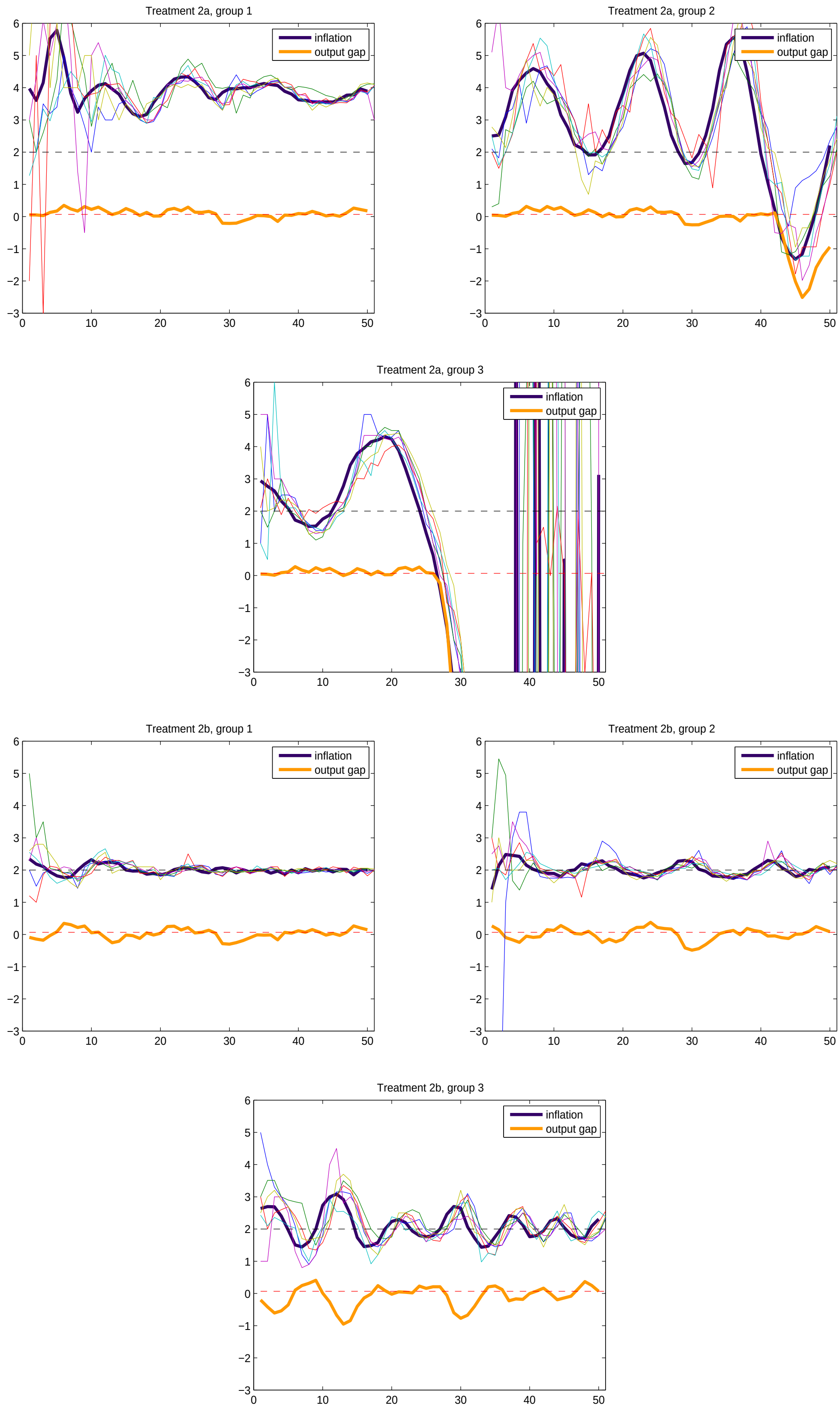


Figure 4: **Upper panels**: Treatment 2a. **Lower panels**: Treatment 2b. Blue thick line: realized inflation; yellow thick line: realized output gap; thin lines: individual forecasts for inflation.

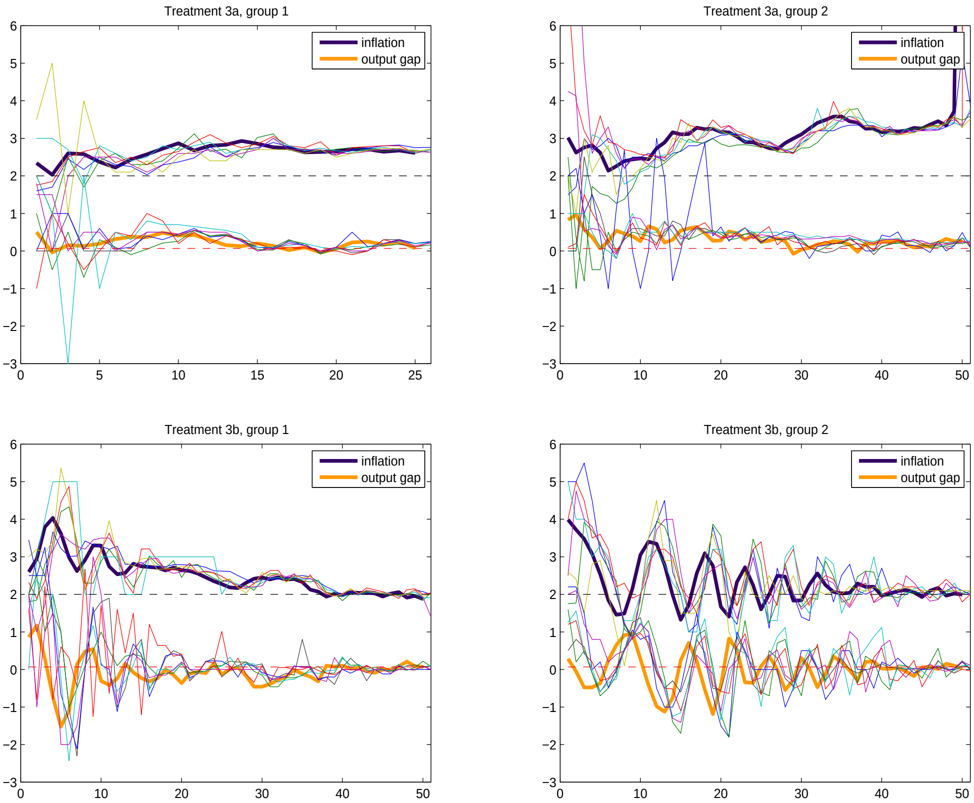


Figure 5: **Upper panels**: Treatment 3a. **Lower panels**: Treatement 3b. Blue thick line: realized inflation; yellow thick line: realized output gap; thin lines: individual forecasts for inflation and output gap.

the lower panels of Fig. 5 we plot the output gap, inflation and individual forecasts for both variables in two sessions of treatment $3b$ ($\phi_\pi = 1.5$). In both groups we observe convergence to the 2 percent fundamental steady state, but the converging paths are different. In group 1, after some initial oscillations, inflation and output gap converge more or less monotonically, while in group 2 the convergence is oscillatory.

Hence, with subjects in the experiment forecasting both inflation and output gap, a monetary policy that responds aggressively to fluctuations in the inflation rate stabilizes fluctuations in inflation and output and leads the economy to the desired outcome.

In order to get more insights into the stabilizing effect of a more aggressive

end effect. In fact, participant 3 predicted an inflation rate of 100% in the last period, causing actual inflation to jump to about 20%.

Group	Inflation	Output gap
1a-1	0.3125	0.0052
1a-2	23.3332	0.0071
1a-3	0.4554	0.0052
1a (median)	0.4554	0.0052
1b-1	0.0715	0.0195
1b-2	0.0169	0.0115
1b-3	0.5100	0.0720
1b (median)	0.0715	0.0195
2a-1	3.8972	0.0181
2a-2	3.7661	0.4953
2a-3	6003.1485	35699.2582
2a (median)	3.8972	0.4953
2b-1	0.0160	0.0265
2b-2	0.0437	0.0400
2b-3	0.1977	0.1383
2b (median)	0.0437	0.0400
3a-1 (excl. t=50)	1.2159	0.1073
3b-1	0.4804	0.1865
3b-2	0.4366	0.2256
3b (median)	0.4585	0.2060

Table 2: Average quadratic difference from the REE

monetary policy, Table 2 summarizes, the quadratic distance of inflation and output gap from its RE fundamental benchmark for all treatments. The table confirms our earlier graphical observation that a more aggressive Taylor rule stabilizes inflation. Increasing the Taylor coefficient from 1 to 1.5 leads to more stable inflation by a factor around 6 in Treatment 1, a factor of 90 in Treatment 2 and a factor of about 3 in Treatment 3. In contrast to inflation the output gap is not stabilized in our experimental economy where the central bank sets the interest rate responding only to inflation.

3 A heterogeneous expectations model

The goal of this section is to characterize individual forecasting behavior and explain the emergence of the three different observed patterns of inflation and output in the experiment, namely convergence to (some) equilibrium level, permanent oscillations and oscillatory convergence, using a simple model of learning.

The fact that different types of aggregate behavior arise in our experiments suggests that heterogeneous expectations play an important role in determining the aggregate outcomes. In fact, a stylized fact that emerged from the investigation of individual experimental data is that there is a pervasive heterogeneity in the forecasting rules used by the subjects in the experiment. We do not report the description of all time series of individual forecasts for the sake of brevity and we refer to Massaro (2012) for a fully detailed analysis of the individual forecasting behaviors observed in the experiment.

Another interesting stylized fact that emerged from the experimental data is that individual forecasting behaviors entail a learning process which takes the form of switching from one heuristic to another. Evidence of switching behavior can be found by inspecting the time series of individual forecasts. Here we report in Fig. 6 some graphical evidence of individual switching behavior.[5]

Fig. 6 shows the time series of some individual forecasts together with the realizations of the variable being forecasted. For every period t we plot the realized inflation or output gap together with the two period ahead forecast of the individual. In this way we can graphically infer how the individual prediction uses the last available observation. For example, if the time series coincide, the subject is using a naive forecasting strategy.

In Fig. 6(a) (group 2, treatment $3a$), subject 2 strongly extrapolates changes in the output gap in the early stage of the experiment, but starting from period $t = 18$ he switches to a much weaker form of trend extrapolation.

In Fig. 6(b) (group 1, treatment $3b$), subject 4 switches between various constant predictors for inflation in the first 23 periods of the experimental session. She is in fact initially experimenting with three predictors, 2% 3% and 5%, and then

[5]Direct evidence of switching behavior has been found in the questionnaires submitted at the end of the experiments, where participants are explicitly asked whether they changed their forecasting strategies throughout the experiment. About 42% of the participants answered that they changed forecasting strategy during the experiment.

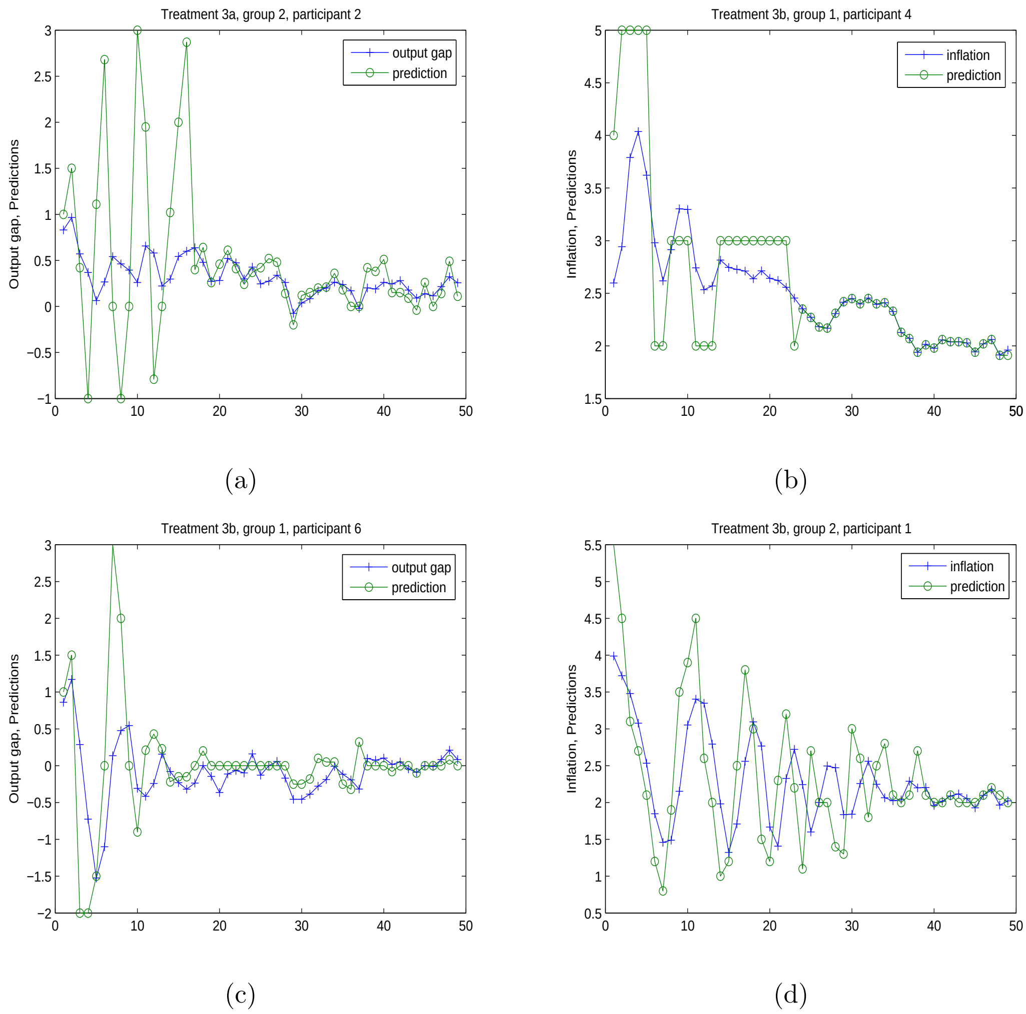


Figure 6: **Individual learning as switching between heuristics.** For every period the subject's forecast $x^e_{i,t+2}$ (green) and the variable being forecast x_t, with $x = \pi, y$, are reproduced.

switches to a naive forecasting strategy after period 23. In the same experimental session, Fig. 6(c), participant 6 predicting the output gap is using different trend extrapolation strategies and, in the time interval $t = 19, ..., 30$, he uses a constant predictor for the output gap. This group illustrates an important point: in the same economy individuals forecasting different variables may use different forecasting strategies.

In Fig. 6(d) group 2, treatment $3b$, subject 1 uses a trend following rule in the initial part of the experiment, i.e. when inflation fluctuates more. However, when oscillations dampen and inflation converges to the equilibrium level, he uses a forecasting strategy very close to naive.

In the light of the empirical evidence for heterogeneous expectations and individual switching behavior, we now introduce a simple model which features evolutionary selection between different forecasting heuristics in order to reproduce individual as well as aggregate experimental data.

Anufriev and Hommes (2012) developed a heuristics switching model along the lines of Brock and Hommes (1997), to explain different price fluctuations in the asset pricing experiment of Hommes, Sonnemans, Tuinstra, and van de Velden (2005). The key idea of the model is that the subjects chose between simple heuristics depending upon their relative past performance. The performance measure of a forecasting heuristic is based on its absolute forecasting error and it has the same functional form as the payoff function used in the experiments. More precisely, the *performance measure* of heuristic h up to (and including) time $t-1$ is given by

$$U_{h,t-1} = \frac{100}{1 + |x_{t-1} - x^e_{h,t-1}|} + \eta U_{h,t-2},$$

with $x = \pi, y$. The parameter $0 \leq \eta \leq 1$ represents the *memory*, measuring the relative weight agents give to past errors of heuristic h.
Given the performance measure, the impact of rule h is updated according to a *discrete choice model with asynchronous updating*

$$n_{h,t} = \delta n_{h,t-1} + (1-\delta)\frac{\exp(\beta U_{h,t-1})}{Z_{t-1}}$$

where $Z_{t-1} = \sum_{h=1}^{H} \exp(\beta U_{h,t-1})$ is a normalization factor. The asynchronous updating parameter $0 \leq \delta \leq 1$ measures the inertia in the impact of rule h, reflecting the fact that not all the participants update their rule in every period or at the same time. The parameter $\beta \geq 0$ represents the intensity of choice measuring how sensitive individuals are to differences in heuristics performances.

The evolutionary model can include an arbitrary set of heuristics. Since our goal is to explain the different observed patterns of inflation and output in the experiment, we keep the number of heuristics as small as possible and consider a model with only four forecasting rules. The heuristics included in the model, summarized in Table 3, were obtained as heuristics describing typical individual forecasting behavior observed and estimated in our macro experiments (see Massaro (2012) for details). Interestingly, the same types of heuristics have been obtained as a description of individual forecasting rules in the learning to forecast experiment of Hommes, Sonnemans, Tuinstra, and van de Velden (2005), and they have been successfully used by Anufriev and Hommes (2012) to explain the different price patterns observed in the asset pricing experiment. Therefore, in order to check the robustness of the heterogeneous expectations model across different settings, we fixed the coefficient values to match exactly the set of heuristics used in Anufriev and Hommes (2012) to explain asset pricing experiments. We then extend the empirical exercise of Anufriev and Hommes (2012) to a multivariate (inflation and output gap) setting, and in treatment 3 we apply the same heuristics switching model to both inflation and output forecasting.

Table 3: Set of heuristics

ADA	adaptive rule	$x^e_{1,t+1} = 0.65x_{t-1} + 0.35x^e_{1,t}$
WTR	weak trend-following rule	$x^e_{2,t+1} = x_{t-1} + 0.4(x_{t-1} - x_{t-2})$
STR	strong trend-following rule	$x^e_{3,t+1} = x_{t-1} + 1.3(x_{t-1} - x_{t-2})$
LAA	anchoring and adjustment rule	$x^e_{4,t+1} = 0.5(x^{av}_{t-1} + x_{t-1}) + (x_{t-1} - x_{t-2})$

In the following sections we proceed with the empirical validation of the evolutionary switching model. We report only simulations for some representative experimental economies (treatment 1b, group 1, treatment 2b, group 3, treatment 3a, group 2, and treatment 3b, group 2) which account for the different aggregate behaviors observed in the experiment. Results for experimental economies with analogous qualitative behavior are similar.

3.1 50-periods ahead simulations

The model is initialized by two initial values for inflation and output gap, π_1, y_1, π_2 and y_2, and initial weights $n_{h,\text{in}}$, $1 \leq h \leq 4$. Given the values of inflation and output gap for periods 1 and 2, the heuristics forecasts can be computed and, using the initial weights of the heuristics, inflation and output gap for period 3, π_3 and y_3, can be computed. Starting from period 4 the evolution according to the model's equations is well defined. Once we fix the four forecasting heuristics, there are three free "learning" parameters left in the model: β, η, and δ. We used the same set of learning parameters as in Anufriev and Hommes (2012), namely $\beta = 0.4$, $\eta = 0.7$, $\delta = 0.9$, and we chose the initial shares of heuristics in such a way to match the patterns observed in the first few periods of the experiment. We also experimented with initial values of inflation and output gap close to the values observed in the first two rounds of the corresponding experimental session. After some trial-and-error experimentation with different initial conditions we were able to replicate all three different qualitative patterns observed in the experiment. For the simulations shown in Fig. 7 we used the same realizations for demand and supply shocks as in the experiment and we chose the initial conditions as follows:

- treatment $1b$, group 1, with convergence to fundamental equilibrium level

  initial inflation rates: $\pi_1 = 2.5$, $\pi_2 = 2.5$;

  initial fractions: $n_{1,\text{in}} = n_{4,\text{in}} = 0.40$, $n_{2,\text{in}} = n_{3,\text{in}} = 0.10$;

- treatment $2b$, group 3, with permanent oscillations

  initial inflation: $\pi_1 = 2.64$, $\pi_2 = 2.70$ (experimental data);

  initial output gap: $y_1 = -0.20$, $y_2 = -0.42$ (experimental data);

  initial fractions: $n_{1,\text{in}} = 0$, $n_{2,\text{in}} = n_{3,\text{in}} = 0.20$, $n_{4,\text{in}} = 0.60$;

- treatment $3a$, group 2, with convergence to a non-fundamental steady state

  initial inflation: $\pi_1 = 2.4$, $\pi_2 = 2.0$;

  initial output gap: $y_1 = 1.8$, $y_2 = 2$;

  initial fractions inflation: $n_{1,\text{in}} = 0.60$, $n_{2,\text{in}} = 0.05$, $n_{3,\text{in}} = 0.10$, $n_{4,\text{in}} = 0.25$

initial fractions output gap: $n_{1,\text{in}} = 0.6$, $n_{2,\text{in}} = 0.05$, $n_{3,\text{in}} = 0.15$, $n_{4,\text{in}} = 0.20$.

- treatment $3b$, group 2, with oscillatory convergence

  initial inflation: $\pi_1 = 3.98$, $\pi_2 = 3.72$ (experimental data);

  initial output gap: $y_1 = 0.28$, $y_2 = -0.05$ (experimental data);

  initial fractions inflation: $n_{1,\text{in}} = 0$, $n_{2,\text{in}} = 0.10$, $n_{3,\text{in}} = 0.40$, $n_{4,\text{in}} = 0.50$

  initial fractions output gap: $n_{1,\text{in}} = 0.15$, $n_{2,\text{in}} = 0.20$, $n_{3,\text{in}} = 0.50$, $n_{4,\text{in}} = 0.15$.

Fig. 7 shows realizations of inflation and output gap in the experiment together with the simulated paths using the heuristics switching model.[6] The model is able to reproduce qualitatively all three different patterns observed in the experiment, which are, convergence to (some) equilibrium, permanent oscillations and oscillatory convergence. As shown in Table 4, the model is also capable to match some quantitative features of the experimental data, such as the mean and the variance.[7]

Table 4: Observed vs simulated moments (50-periods ahead)

	Treatment											
	1b		2b		3a (π)		3a (y)		3b (π)		3b (y)	
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
Obs.	2.19	0.01	2.05	0.18	3.06	0.14	0.29	0.03	2.20	0.23	-0.02	0.23
Sim.	2.15	0.02	2.03	0.17	3.13	0.05	0.24	0.04	2.03	0.22	-0.07	0.33
p	0.06	0.03	0.74	0.71	$*$	$*$	0.32	0.19	0.01	0.67	0.50	0.05

$*$ = Non stationarity.
The row corresponding to p reports p-values of tests on the equality of observed and simulated mean and on the equality of observed and simulated variance (HAC Consistent covariance estimators (Newey-West) have been used to compute standard errors).

[6] Treatment 3a group 2 has been simulated for 49 periods due to a clear ending effect, see footnote 4.

[7] We performed the tests on the equality of observed and simulated mean and variance on a sample that goes from period 4 to the end of the experimental session in order to minimize the impact of the initial conditions.

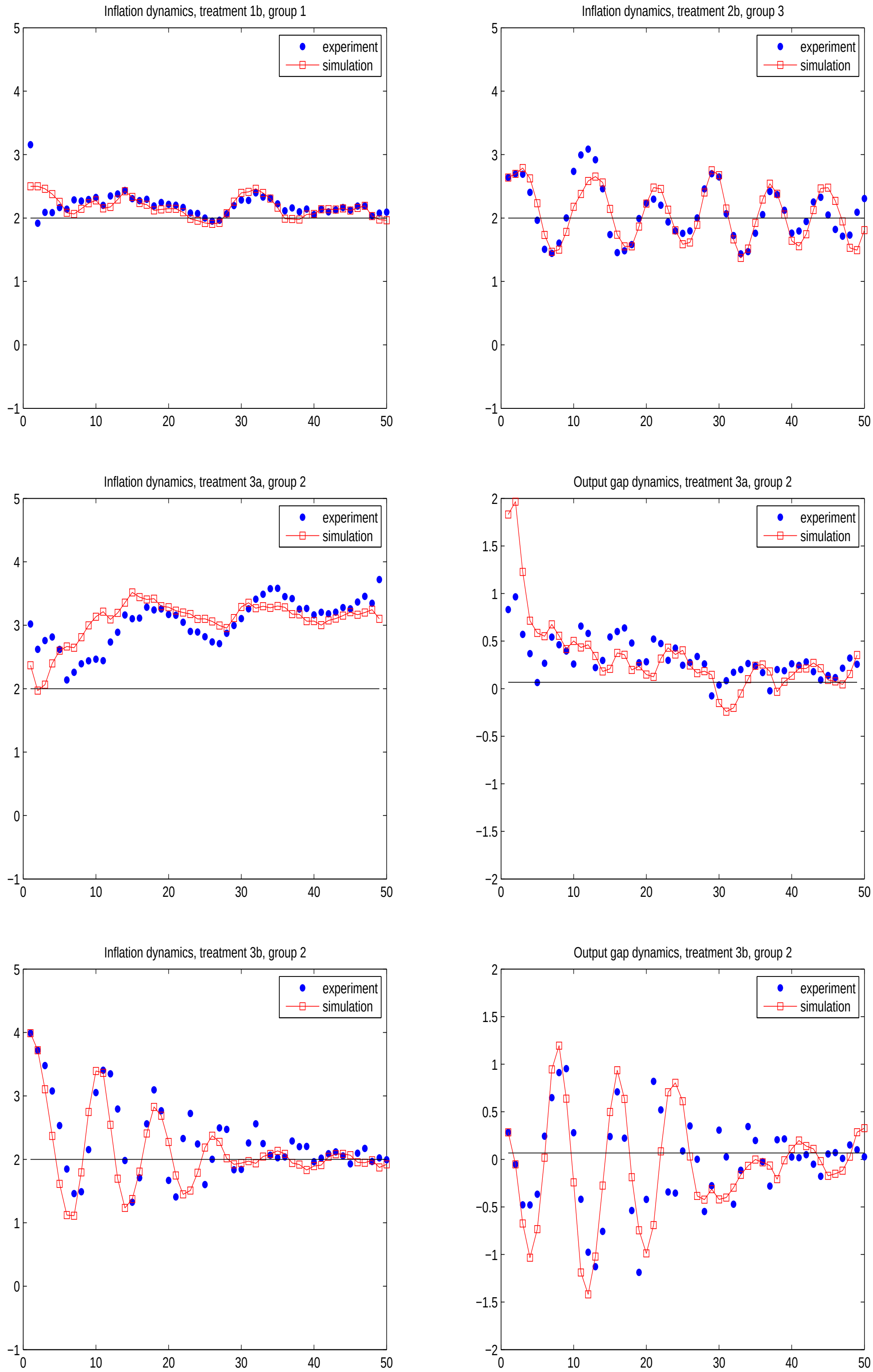


Figure 7: Experimental data (blue points) and **50-periods ahead** heuristics switching model simulations (red lines)

3.2 One-period ahead simulations

The 50-period ahead simulations fix initial states and then predicts inflation and output patterns 50-periods ahead. We now report the results of one-step ahead simulations of the nonlinear switching model. At each time step, the simulated path uses *experimental data* as inputs to compute the heuristics' forecasts and update their impacts. Hence, the one-period ahead simulations use exactly the same information as the subjects in the experiments. The one-period ahead simulations match the different patterns in the experimental data quite nicely. Fig. 8 compares the experimental data with the one-step ahead predictions made by our model, using the benchmark parameter values $\beta = 0.4$, $\eta = 0.7$, $\delta = 0.9$. In these simulations initial inflation and output gap initial inflation and output gap in the first two periods are taken from the corresponding experimental group, while the initial impacts of all heuristics are equal to 0.25.

Fig. 9 shows how in different groups different heuristics are taking the lead after starting from a uniform distribution. In treatment $1b$ group 1 (Fig. 9(a)), the initial drop in inflation, from 3.1 to 1.9 respectively in periods 1 and 2, causes an overshooting in the predictions of the trend extrapolating rules, i.e. WTF, STF and LAA, for inflation in period 3. Therefore the relative impacts of these rules starts to drop, while the relative share of adaptive expectations ADA increases to about 70% in the first 14 periods. From period 14 on, the share of the WTF rule increases due to some slow oscillation, and it reaches a peak of about 48% in period 33. During this time span of slow oscillations the fraction of the ADA rule decreases to about 30%. However, in the last part of the experiment inflation stabilizes and the ADA rule dominates the other rules. In group 3 treatment $2b$ (Fig. 9(b)) we clearly observe that the ADA rule is not able to match the oscillatory pattern and its impact declines monotonically in the simulation. The STF rule can follow the oscillatory pattern and initially dominates (almost 40% in period 8) but its predictions overshoot the trend in realized inflation reverses, and its relative share declines monotonically from period 9 on. Both the WTF and the

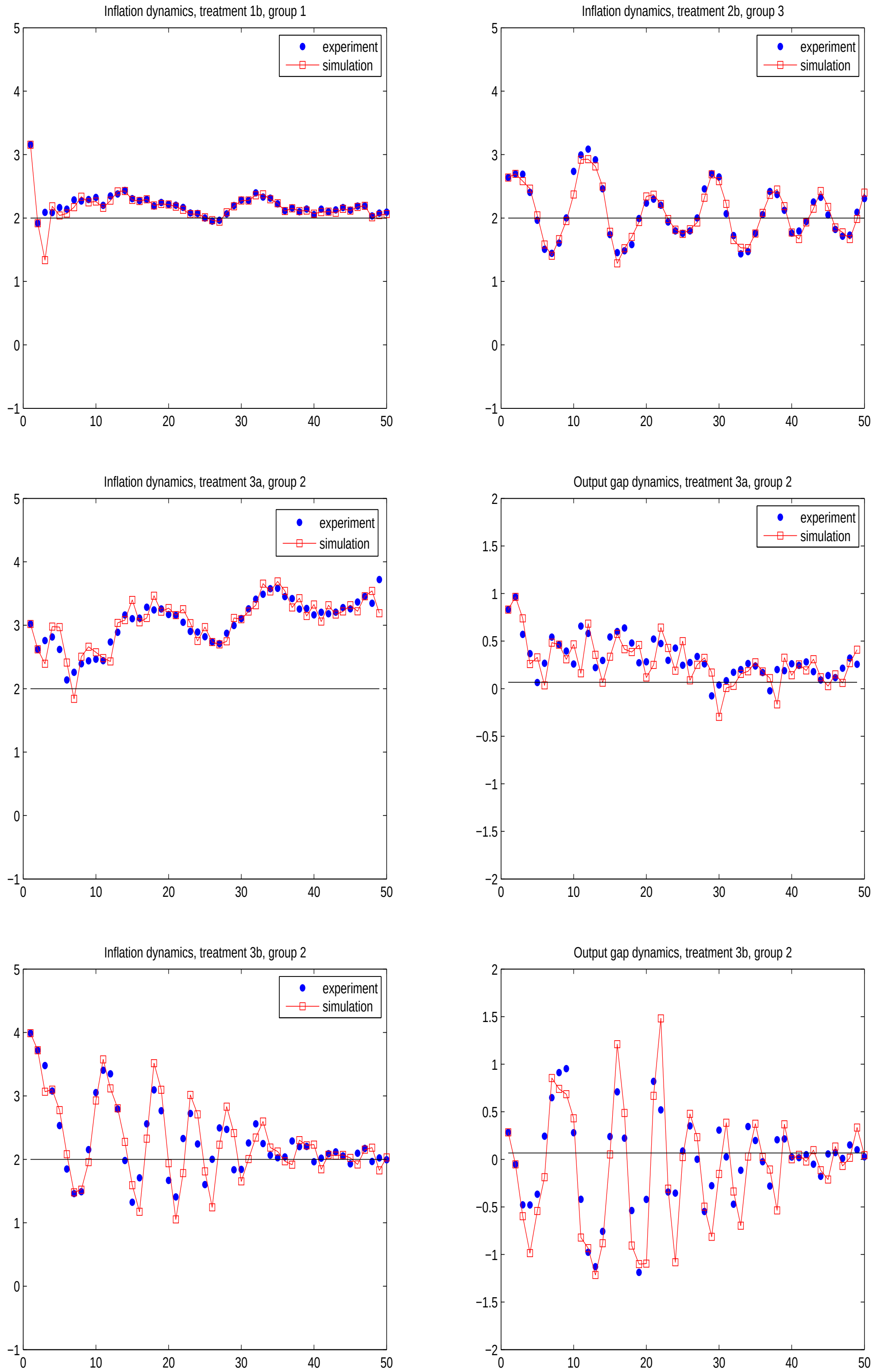


Figure 8: Experimental data (blue points) and **one-period ahead** heuristics switching model simulations (red lines)

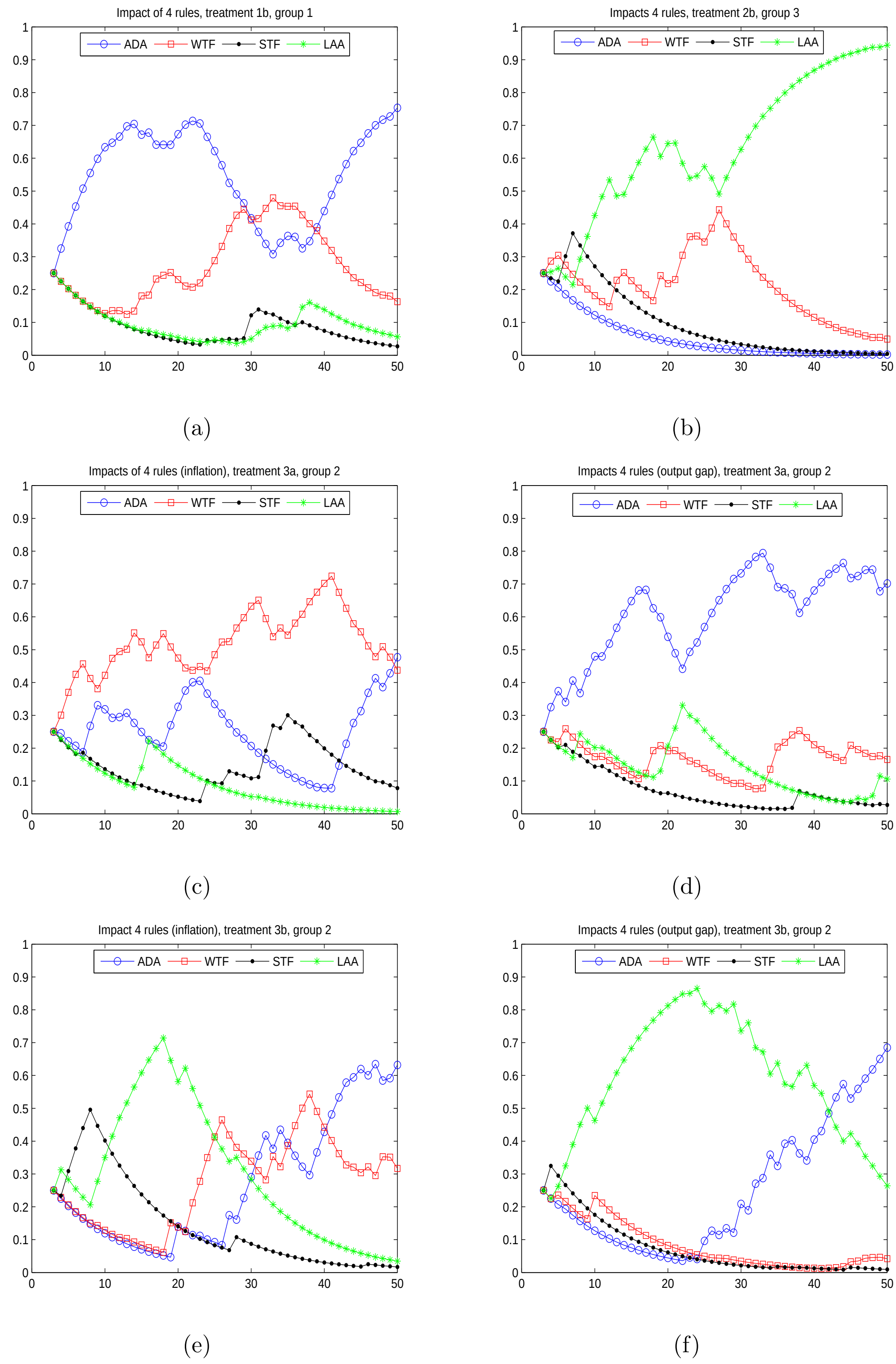


Figure 9: Evolution of fractions of 4 heuristics corresponding to one-period ahead simulations in Fig. 8: adaptive expectations (ADA, blue), weak trend follower (WTF, red), strong trend follower (STF, black), anchoring and adjustment heuristics (LAA, green).

LAA rule can follow closely the observed oscillations, but in the last part of the experiment the LAA rule dominates the other rules. As in the quite different setting of the asset pricing experiments in Anufriev and Hommes (2012), our simulation explains oscillatory behavior by coordination on the LAA rule by most subjects. In the early stage of treatment $3a$, group 2 (Fig. 9(c)), the oscillations in inflation are relatively small and therefore the WTF rule is able to match the oscillatory pattern; also the ADA rule performs reasonably well, while both the STF and LAA rules overshoot too often. Then inflation undergoes a more turbulent phase with stronger oscillations starting in period 24 and the impact of the strong trend following rule increases and reaches a peak of about 30% in period 35. At the same time, when inflation fluctuates the share of the ADA rule declines. In the last part of the experiment inflation more or less stabilizes and the impact of the WTF rule declines monotonically, while, the impact of the ADA rule rises from less than 10% to about 50% in the last 10 periods of the experiment. Interestingly, in the same economy the story is different for the output gap (9(d)). In fact the dynamics are characterized by oscillations in the early stage of the experiment which are less pronounced than the oscillations in the inflation rate. The model then explains the convergence pattern of output gap with small oscillation by coordination of most individuals on the ADA rule and a share of WTF that varies between 7% and 25% throughout the experiment. A novel feature of our heuristics switching model is that it allows for coordination on different forecasting rules for different aggregate variable of the same economy. Inflation expectations are dominated by weak trend followers, causing inflation to slowly drift away to the "wrong" non-fundamental steady state, while output expectations are dominated by adaptive expectations, causing output to converge (slowly) to its fundamental steady state level.

For treatment $3b$ group 2 (Fig. 9(e)), the one step ahead forecast exercise produces a rich evolutionary competition among heuristics. In the initial part of the experiment, the STF is the only rule able to match the strong decline in the inflation rate and its share increases to 50% in period 8. However the impact of the

STF rule starts to decrease after it misses the first turning point. After the initial phase of strong trend in inflation, the LAA rule does a better job in predicting the trend reversal and its impact starts to increase, reaching a share of about 70% in period 18. However oscillations slowly dampen and therefore the impacts of the ADA rule and the WTF rule starts to rise. Towards the end of the simulation, when inflation has converged, the ADA rule dominates the other heuristics. The evolutionary selection dynamics are somewhat different for the output gap predictors (Fig. 9(f)). In fact, oscillations of the output gap are more frequent and this implies a relatively bad forecasting performance of the STF rule that tends to overshoot more often. The switching model explains the oscillatory behavior of output in the initial phase by coordination on the LAA rule by most subjects. However, with dampening oscillations the impact of the LAA rule gradually decreases and the ADA rule starts increasing after period 25 and dominates in the last 10 periods. Fig. 10 reports the predictions of the participants in the experiments together with the predictions generated by the four heuristics, while Table 5 compares observed and simulated moments.[8]

Table 5: Observed vs simulated moments (one-period ahead)

	Treatment											
	1b		2b		3a (π)		3a (y)		3b (π)		3b (y)	
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
Obs.	2.19	0.01	2.05	0.18	3.06	0.14	0.29	0.03	2.20	0.23	-0.02	0.23
Sim.	2.17	0.01	2.05	0.16	3.08	0.14	0.25	0.04	2.21	0.32	-0.09	0.40
p	0.02	0.86	0.83	0.16	0.24	0.78	0.01	0.15	0.60	0.05	0.03	0.01

The row corresponding to p reports p-values of tests on the equality of observed and simulated mean and on the equality of observed and simulated variance (HAC Consistent covariance estimators (Newey-West) have been used to compute standard errors).

[8]We performed tests on the equality of observed and simulated mean and variance on a sample that goes from period 4 to the end of the experimental session in order to minimize the impact of initial conditions.

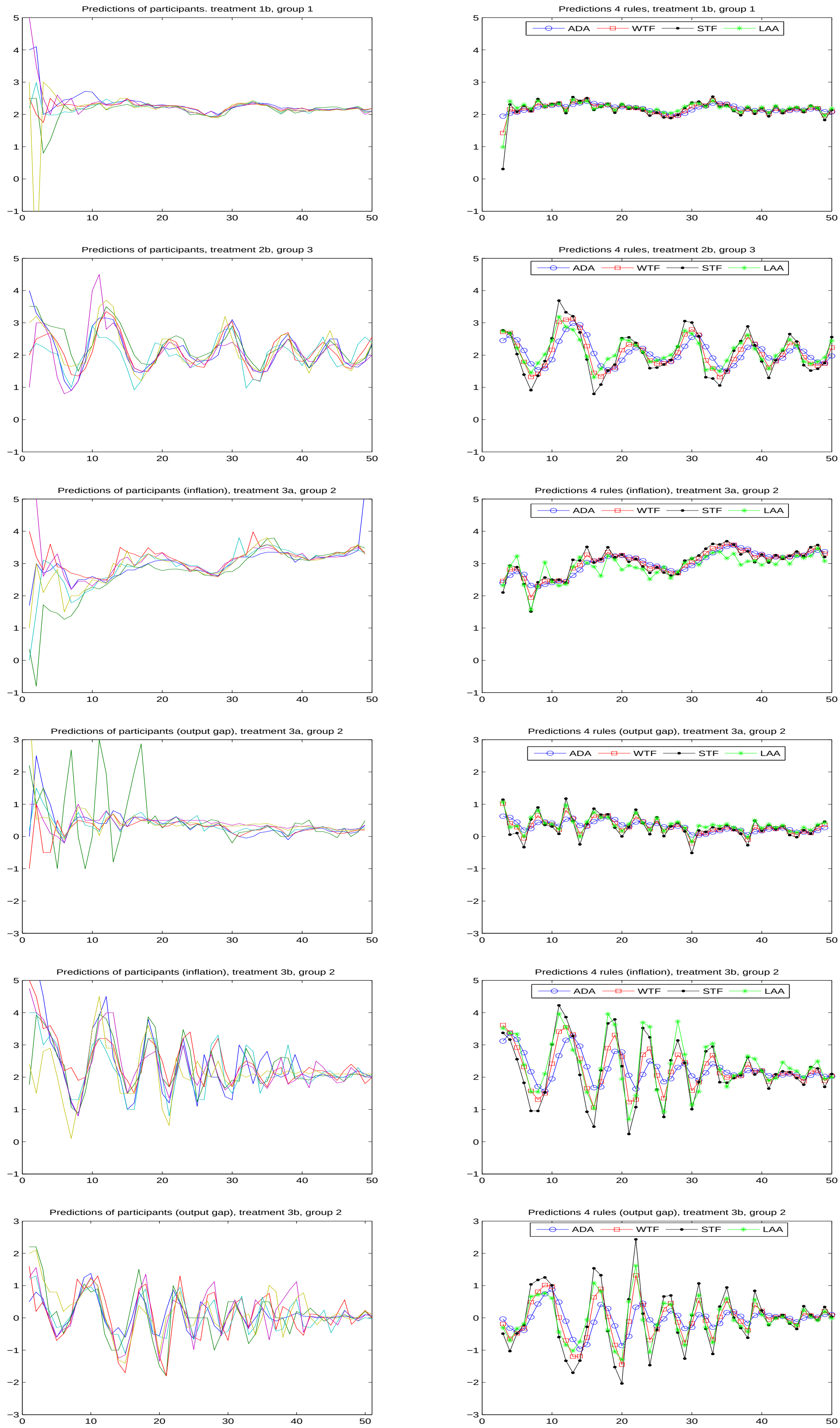


Figure 10: **Left panels:** predictions of the participants in the experiment. **Right panels:** predictions of the four heuristics.

Forecasting performance

Table 6 compares the MSE of the one-step ahead prediction in 10 experimental groups[9] for 9 different models: the rational expectation prediction (RE), six homogeneous expectations models (naive expectations, fixed anchor and adjustment (AA) rule,[10] and each of the four heuristics of the switching model), the switching model with benchmark parameters $\beta = 0.4$, $\eta = 0.7$, and $\delta = 0.9$, and the "best" switching model fitted by means of a grid search in the parameters space. The MSEs for the benchmark switching model are shown in bold and, for comparison, for each group the MSEs for the best among the four heuristics are also shown in bold. The best among all models for each group is shown in italic.[11] We notice immediately that the RE prediction is (almost) always the worst. It also appears that the evolutionary learning model is able to make the best out of different heuristics. In fact, none of the homogeneous expectations models fits *all* different observed patterns, while the best fit switching model yields the lowest MSE in 9/15 cases,[12] being the second best, with only a slightly larger MSE compared to the best model, in the other cases (with the exceptions of group 1 in treatment $2a$ and groups 2 and 3 in treatment $3a$). Notice also that the benchmark switching model typically is almost as good as the best switching model, indicating that the results are not very sensitive to the learning parameters.

[9] The MSE of the one-step ahead prediction for the remaining groups is reported in the Supplementary material, Table 11

[10] In the AA rule we consider the full sample mean, which is a proxy of the equilibrium level, as an anchor. In the LAA rule instead we use the sample average of all the previous realizations that are available at every point in time as an anchor.

[11] We evaluate the MSE over 47 periods, for $t = 4, ..., 50$. This minimizes the impact of initial conditions for the switching model in the sense that $t = 4$ is the first period when the prediction is computed with both the heuristics forecasts and the heuristics impacts being updated on the basis of experimental data.

[12] We excluded treatment $1a$, group 2 and did not fit the heuristics switching model because of the anomalous observed behavior.

Table 6: MSE over periods $4-50$ of the one-period ahead forecast.

Model	**Tr1*a* gr1**	**Tr1*b* gr1**	**Tr1*b* gr3**	**Tr2*a* gr1**	**Tr2*a* gr2**	**Tr2*b* gr1**	**Tr2*b* gr3**	**Tr3*a* gr2**	**Tr3*b* gr1**	**Tr3*b* gr2**
RE	0.3366	0.0438	0.5425	Indeterminacy		0.0122	0.1822	Indeterminacy	0.5816	0.4851
naive	0.0058	*0.0016*	0.0454	0.1253	0.5024	0.0032	0.0552	0.0579	0.1126	0.2855
AA	0.0181	0.0066	0.1273	*0.0487*	0.4099	0.0066	0.0185	0.1170	0.1533	*0.1746*
ADA	0.0098	**0.0021**	0.0908	0.3170	0.9893	0.0110	0.1113	***0.0531***	0.1536	0.3881
WTF	**0.0045**	0.0026	0.0209	0.0840	0.2652	**0.0035**	0.0273	0.0705	**0.1060**	0.2215
STF	0.0137	0.0106	**0.0084**	0.1359	**0.1749**	0.0131	0.0474	0.1383	0.2266	0.4329
LAA	0.0191	0.0066	0.1243	**0.0606**	0.3931	0.0064	**0.0147**	0.0985	0.1302	**0.1870**
4 rules (benchmark)	**0.0048**	**0.0017**	**0.0117**	**0.0692**	**0.0934**	**0.0024**	**0.0092**	**0.0632**	**0.0958**	**0.1898**
4 rules (best fit)	*0.0044*	*0.0016*	*0.0089*	0.0665	*0.0897*	*0.0022*	0.0093	0.0609	*0.0936*	0.1840
β	1	10	2	0	10	10	1	10	7	1
η	0.9	0.7	0.7	0	0.4	0.1	0.7	0.9	0.1	0.5
δ	0.5	0.8	0.7	0	0.9	0.8	0.9	0.8	0.8	0.9

Note that the MSE reported for treatment 3 refers to the sum of the MSE relative to inflation and the MSE relative to output gap.
In treatment 3a, group 2, the MSE has been computed for periods $4-49$ due to the observed ending effect.
Moreover the fact that in treatment 2b, group 3, the MSE reported for the benchmark model is lower than the MSE of the best fit model is due to the fact that the grid search for parameter β takes a step of 1 and thus excludes $\beta = 0.4$.

Out-of-sample forecasting

In order to evaluate the out-of-sample forecasting performance of the model, we first perform a grid search to find the parameters of the model minimizing the MSE for a restricted sample, i.e. for periods $t = 4, ..., 43$. Then, the squared forecasting errors are computed for the next 7 periods. The results are shown in Table 7 and in the Supplementary material, Table 12. Finally, we compare the out-of-sample forecasting performance of the structural heuristics switching model (both the best fit and the model with benchmark parameters) with a simple non-structural AR(2) model with three parameters. Notice that, for treatment 3 we use different AR(2) models for inflation and output gap, so that we have in fact 6 parameters for the AR(2) models in treatment 3.

For the converging groups (treatment $1a$ groups 1 and 3, treatment $1b$ groups 1 and 2, treatment $2a$ group 1, treatment $2b$ groups 1 and 2, treatment $3a$ groups 1 and 2, treatment $3b$ group 1) we typically observe that the squared prediction errors remain very low and comparable with the MSEs computed in-sample. This is due to the fact that the qualitative behavior of the data does not change in the last periods. For the groups that exhibit oscillatory behavior (treatment $1b$ group 3, treatment $2a$, groups 2 and 3, treatment $2b$ group 3) the out-of-sample errors are larger than the in-sample MSEs, and they typically increase with the time horizon of the prediction. When we instead observe dampening oscillations (treatment $3b$, group 2), the out-of-sample prediction errors are smaller than the in-sample MSEs. This is due to the fact that, towards the end of the experimental session, convergence is observed. Comparing the out-of-sample forecasting performance, we conclude that the benchmark switching model generally does not perform worse (sometimes even better) than the best in-sample fitted switching model. Compared to the non-structural AR(2) model, the switching model on average performs better. In particular, for treatment 3 the benchmark switching model as well as the 3-parameter best-fit switching model perform better than the AR(2) models with 6 parameters.

Table 7: Out-of-sample performance.

	Tr1*a* gr1	**Tr1*b* gr1**	**Tr1*b* gr3**	**Tr2*a* gr1**	**Tr2*a* gr2**	**Tr2*b* gr1**	**Tr2*b* gr3**	**Tr3*a* gr2**	**Tr3*b* gr1**	**Tr3*b* gr2**
Best Fit Switching Model										
(β, η, δ)	(1,0.9,0.5)	(10,0.7,0.8)	(3,0.7,0.5)	(0,0,0)	(10,0.4,0.9)	(10,0.1,0.8)	(1,0.7,0.9)	(10,0.9,0.8)	(5,0.1,0.8)	(2,0.5,0.9)
1 p ahead	0.0000	0.0000	0.0287	0.0006	0.1201	0.0000	0.0007	0.0060	0.0113	0.0245
2 p ahead	0.0020	0.0002	0.0046	0.0003	0.0511	0.0037	0.1076	0.0085	0.0170	0.0501
3 p ahead	0.0011	0.0000	0.0256	0.0006	0.0931	0.0014	0.2364	0.0033	0.0059	0.0721
4 p ahead	0.0016	0.0000	0.1930	0.0006	0.0347	0.0038	0.1902	0.0224	0.0344	0.0663
5 p ahead	0.0083	0.0005	0.4156	0.0070	0.0131	0.0022	0.0056	0.0611	0.0669	0.0131
6 p ahead	0.0004	0.0020	0.6021	0.0690	0.4067	0.0023	0.1732	0.0259	0.0125	0.0491
7 p ahead	0.0003	0.0041	0.6235	0.0271	1.5804	0.0007	0.3218	0.3048	0.0061	0.0727
Benchmark Switching Model										
(β, η, δ)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)
1 p ahead	0.0000	0.0004	0.0388	0.0015	0.1087	0.0005	0.0042	0.0062	0.0229	0.0386
2 p ahead	0.0009	0.0003	0.0382	0.0016	0.1442	0.0004	0.0957	0.0045	0.0062	0.0442
3 p ahead	0.0004	0.0008	0.0190	0.0028	0.5843	0.0002	0.1404	0.0051	0.0085	0.0343
4 p ahead	0.0000	0.0005	0.0039	0.0026	1.1437	0.0014	0.0659	0.0134	0.0238	0.0337
5 p ahead	0.0004	0.0017	0.0028	0.0042	1.4969	0.0008	0.0080	0.0464	0.0363	0.0284
6 p ahead	0.0000	0.0049	0.0019	0.0589	0.9876	0.0021	0.2232	0.0271	0.0393	0.0361
7 p ahead	0.0000	0.0075	0.0198	0.0235	0.1579	0.0007	0.1997	0.2575	0.0181	0.0240
AR(2) Model										
$(\beta_{\pi,0}, \beta_{\pi,1}, \beta_{\pi,2})$	$(0.3, 1.1, -0.2)$	$(0.4, 0.7, 0.1)$	$(0.2, 1.8, -0.9)$	(1.8,1.3,-0.7)	(0.6,1.8,-1.0)	(0.7,0.9,-0.2)	(0.9,1.4,-0.9)	(0.3,1.1,-0.2)	(0.5,1.3,-0.5)	(1.3,1.1,-0.6)
$(\beta_{y,0}, \beta_{y,1}, \beta_{y,2})$								(0.2,0.6,-0.2)	(-0.1,0.9,-0.7)	(-0.0,0.9,-0.7)
1 p ahead	0.0034	0.0007	0.0182	0.0288	0.0749	0.0001	0.0155	0.0166	0.0212	0.0343
2 p ahead	0.0005	0.0007	0.0278	0.1631	1.0720	0.0045	0.1795	0.0612	0.1648	0.1849
3 p ahead	0.0146	0.0007	0.0019	0.2238	4.2891	0.0004	0.2624	0.0407	0.1723	0.0868
4 p ahead	0.0347	0.0005	0.0190	0.1600	9.4516	0.0003	0.1573	0.0859	0.2195	0.0307
5 p ahead	0.0000	0.0202	0.0108	0.1081	17.480	0.0215	0.0331	0.1096	0.4766	0.1323
6 p ahead	0.0007	0.0105	0.0339	0.0016	23.415	0.0002	0.0741	0.0455	0.4003	0.0735
7 p ahead	0.0009	0.0085	0.0756	0.0005	24.565	0.0000	0.2055	0.3515	0.5263	0.0729

Note that the squared prediction errors for treatment 3 refer to the sum of the errors relative to inflation and to output gap.
In treatment $3a$, group 2, the switching models and the AR(2) model have been estimated on a restricted sample of 42 periods due to the observed ending effect.

4 Conclusions

In this paper we use laboratory experiments with human subjects to study individual expectations, their interactions and the aggregate behavior they co-create within a New Keynesian macroeconomic setup and we fit a heterogeneous expectations switching model to the experimental data. A novel feature of our experimental design is that realizations of aggregate variables depend on individual forecasts of two different variables, the output gap and inflation. We find that individuals tend to base their predictions on past observations, following simple forecasting heuristics, and individual learning takes the form of switching from one heuristic to another. We propose a simple model of evolutionary selection among forecasting rules based on past performance in order to explain individual forecasting behavior as well as the different aggregate outcomes observed in the laboratory experiments, namely convergence to some equilibrium level, persistent oscillatory behavior and oscillatory convergence. Our model is the first to describe aggregate behavior in a stylized macro economy as well as individual micro behavior of heterogeneous expectations about two different variables. A distinguishing feature of our heterogeneous expectations model is that evolutionary selection may lead to different dominating forecasting rules for different variables within the same economy, for example a weak trend following rule dominates inflation forecasting while adaptive expectations dominate output forecasting (see Figs. 9(c) and 9(d)).

We also perform an exercise of empirical validation on the experimental data to test the model's performance in terms of in-sample forecasting as well as out-of-sample predicting power. Our results show that the heterogeneous expectations model outperforms models with homogeneous expectations, including the rational expectations benchmark.

On the policy side we find that the implementation of a monetary policy that reacts aggressively to deviations of inflation from the target leads the economy to the desired outcome, but only in the (very) long run for the more relevant treatment 3. The convergence to the desired target can be slow, e.g., more than 20 periods (quarters) as in the case of treatment 3 where subjects forecast inflation and the output gap. This is due to the fact that, in contrast to standard RE models that display weak internal propagation in response to shocks (see e.g., Chari, Kehoe, and McGrattan (2000) and Nelson (1998)), a simple model with heterogeneous expectations and evolutionary switching can generate persistent deviations from steady state. As an example, consider Fig. 11 which shows the responses of

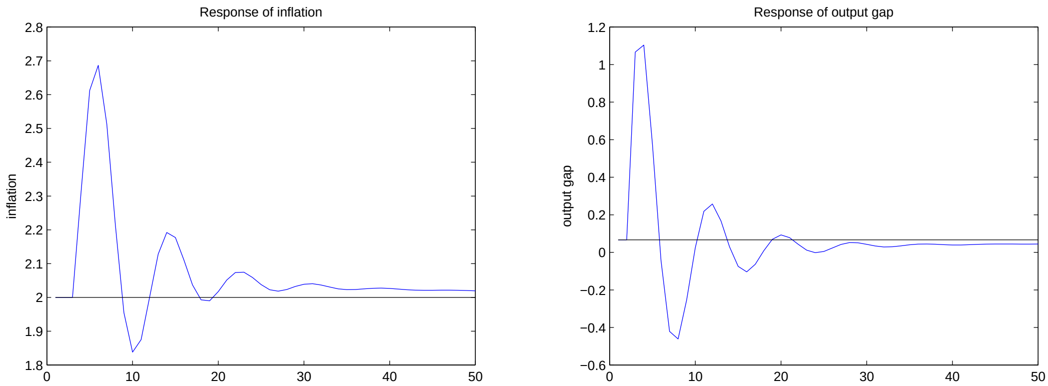


Figure 11: Impulse responses to one unit demand shock

inflation and the output gap to a demand shock within a NK environment where rational expectations are replaced by our switching model with uniform initial distribution of heuristics (as in our simulations of treatment 3). In the homogeneous RE benchmark, after the shock, the system immediately jumps back to the desired equilibrium level. In contrast, Fig. 11 shows that our heterogeneous expectations model exhibits slow convergence to the RE steady state and persistent fluctuations in response to shocks, even in the presence of an aggressive monetary policy ($\phi_\pi = 1.5$). In fact, while the system converges in the long run, in the short run oscillations in inflation and output arise due to coordination of individual expectations on trend following rules. These findings are in line with the theoretical results in the monetary policy literature that an interest rate rule following the Taylor principle leads the economy to the desired target, but convergence can be very slow, confirming thus the findings of Sargent (1999), Evans and Honkapohja

(2001), Adam (2005) and Milani (2007), among others, who showed that simple economic models display strong internal propagation once the assumption of rational expectations is relaxed. The value added by our paper is that these results are extended to the (realistic) case of *heterogeneous expectations* both in a laboratory setting and in a theoretical heuristic switching model.

These results stress the importance of the empirical validation of heterogeneous agents models on real macroeconomic data. See e.g. Cornea, Hommes, and Massaro (2012), fitting a 2-type switching model between fundamental and naive expectations to inflation data.

References

Adam, K. (2005): "Learning to forecast and cyclical behavior of output and inflation," *Macroeconomic Dynamics*, 9(1), 1–27.

——— (2007): "Experimental Evidence on the Persistence of Output and Inflation," *The Economic Journal*, 117(520), 603–636.

Anufriev, M., T. Assenza, C. H. Hommes, and D. Massaro (2013): "Interest Rate Rules and Macroeconomic Stability under Heterogeneous Expectations," *Macroeconomic Dynamics*, forthcoming.

Anufriev, M., and C. H. Hommes (2012): "Evolutionary Selection of Individual Expectations and Aggregate Outcomes," *American Economic Journal: Microeconomics*, forthcoming.

Bao, M., C. Hommes, J. Sonnemans, and J. Tuinstra (2012): "Individual Expectations, Limited Rationality and Aggregate Outcomes," *Journal of Economic Dynamics and Control*, 36, 1101–1120.

Branch, W. (2004): "The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations," *The Economic Journal*, 114(497), 592–621.

Branch, W., and G. Evans (2006): "Intrinsic Heterogeneity in Expectation Formation," *Journal of Economic Theory*, 127, 264–295.

Branch, W., and B. McGough (2009): "A new Keynesian Model with Heterogeneous Expectations," *Journal of Economic Dynamics and Control*, 33(5), 1036–1051.

——— (2010): "Dynamic Predictors Selection in a New Keynesian Model with Heterogeneous Expectations," *Journal of Economic Dynamics and Control*, 34(8), 1492–1508.

Brock, W. A., and C. H. Hommes (1997): "A Rational Route to Randomness," *Econometrica*, 65(5), 1059–1095.

Bullard, J., and K. Mitra (2002): "Learning about monetary policy rules," *Journal of Monetary Economics*, 49(6), 1105–1129.

Capistran, C., and A. Timmermann (2009): "Disagreement and biases in inflation expectations," *Journal of Money, Credit and Banking*, 41, 365–396.

Chari, V., P. Kehoe, and E. McGrattan (2000): "Sticky price models of the business cycle: Can the contract multiplier solve the persistence problem?," *Econometrica*, 68, 1151–1179.

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 113.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115(1), 147–180.

Cornea, A., C. Hommes, and D. Massaro (2012): "Behavioral Heterogeneity in U.S. Inflation Dynamics," mimeo, University of Amsterdam.

De Grauwe, P. (2010): "Animal Spirits and Monetary Policy," *Economic Theory*.

Duffy, J. (2008): "Experimental Macroeconomics," in *New Palgrave Dictionary of Economics*, ed. by S. Durlauf, and L. Blume. New York: Palgrave Macmillan, forthcoming.

Evans, G., and S. Honkapohja (1998): "Economic Dynamics With Learning: New Stability Results," *Review of Economic Studies*, 65, 23–44.

——— (2003): "Expectations and the Stability Problem for Optimal Monetary Policies," *Review of Economic Studies*, 70, 807–824.

Evans, G. W., and S. Honkapohja (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.

Gali, J. (2008): *Monetary Policy, Inflation and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.

Gali, J., and M. Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, 44, 195–222.

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., J. Sonnemans, J. Tuinstra, and H. van de Velden (2005): "Coordination of Expectations in Asset Pricing Experiments," *Review of Financial Studies*, 18(3), 955–980.

Lucas, R. E. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4, 103–124.

Mankiw, N., R. Reis, and J. Wolfers (2003): "Disagreement about inflation expectations," Discussion paper, NBER.

Marimon, R., S. E. Spear, and S. Sunder (1993): "Expectationally Driven Market Volatility: An Experimental Study," *Journal of Economic Theory*, 61(1), 74–103.

Marimon, R., and S. Sunder (1994): "Expectations and Learning Under Alternative Monetary Regimes: An Experimental Study," *Economic Theory*, 4, 131–162.

Massaro, D. (2012): "Bounded Rationality and Heterogeneous Expectations in Macroeconomics," Phd thesis, University of Amsterdam.

Milani, F. (2007): "Expectations, Learning and Macroeconomic Persistence," *Journal of Monetary Economics*, 54, 2065–2082.

Muth, J. F. (1961): "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29(3), 315–335.

Nelson, E. (1998): "Sluggish inflation and optimization models of the business cycle," *Journal of Monetary Economics*, 42, 303–322.

Pfajfar, D., and E. Santoro (2010): "Heterogeneity, learning and inflation stickiness in inflation expectations," *Journal of Economic Behavior and Organization*, 75(3), 426–444.

Pfajfar, D., and B. Zakelj (2010): "Inflation Expectations and Monetary Policy Design: Evidence from the Laboratory," mimeo, Tilburg University.

Preston, B. (2005): "Learning about Monetary Policy Rules when Long-Horizon Forecasts Matter," *International Journal of Central Banking*, 1(2), 81–126.

Sargent, T. J. (1999): *The Conquest of American Inflation*. Princeton University Press.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycle: a Bayesian DSGE Approach," *American Economic Review*, 97(3), 586–607.

Woodford, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

SUPPLEMENTARY MATERIAL

- **(Translation of Dutch) Instructions for participants**

- **MSE of the one-period ahead forecast**
  - Table 8

- **Out of sample forecasting performance of the switching model**
  - Table 9

(Translation of Dutch)Instructions for participants (inflation forecasters)

Set-up of the experiment

You are participating in an experiment on economic decision-making. You will be rewarded based on the decisions you make during the experiment. The experiment will be preceded by several pages of instructions that will explain how it works. When the experiment has ended, you will be asked to answer some questions about how it went.

- The whole experiment, including the instructions and the questionnaire, is computerized. Therefore you do not have to submit the paper on your desk. Instead, you can use it to make notes.
- There is a calculator on your desk. If necessary, you can use it during the experiment.
- If you have any question during the experiment, please raise your hand, then someone will come to assist you.

General information about the experiment

In the experiment, *statistical research bureaus* make predictions about the *inflation* and the so-called *"output gap"* in the economy. A limited amount of research bureaus is active in the economy. You are a research bureau that makes predictions about *inflation.* This experiment consists of *50 periods* in total. In each period you will be asked to predict the inflation; your *reward after the experiment has ended* is based on the accuracy of your predictions.

In the following instructions you will get more information about the *economy* you are in, about the way in which *making predictions* works during the experiment, and about the way in which your *reward* is calculated. Also, the *computer program* used during the experiment will be explained.

Information about the economy (part 1 of 2)

The economy you are participating in is described by *three variables*: the inflation π_t, the output gap y_t and the interest rate i_t. The subscript t indicates the *period* the experiment is in. In total there are *50 periods*, so t increases during the experiment from 1 through 50.

The *inflation* measures the percentage change in the price level of the economy. In each period, inflation depends on the *inflation predictions* and *output gap predictions* of the statistical research bureaus, and on *minor price shocks.* There is a *positive relation* between the actual inflation and both the inflation predictions and output gap predictions of the research bureaus. This means for example that if the inflation prediction of a research bureaus increases, then actual inflation will also increase (assuming that the other predictions and the price shock remain equal). The minor price shocks have an equal chance of influencing inflation positively or negatively.

Information about the economy (part 2 of 2)

The *output gap* measures the percentage difference between the *Gross Domestic Product (GDP)* and the *natural GDP*. The *GDP* is the value of all goods produced during a period in the economy. The *natural GDP* is the value the total production would have if prices in the economy would be fully flexible. If the output gap is positive (negative), the economy therefore produces more (less) than the natural GDP. In each period the output gap depends on the *inflation predictions* and *output gap predictions* of the statistical bureaus, on the *interest rate* and on *minor economic shocks.* There is a *positive relation* between the output gap and the inflation predictions and output gap predictions, and a *negative relation* between the output gap and the interest rate. The minor economic shocks have an equal chance of influencing the output gap positively or negatively.

The *interest rate* measures the price of borrowing money and is determined by the *central bank*. There is a *positive relation* between the interest rate and the inflation.

Information about making predictions

Your task, in each period of the experiment, consists in predicting the *inflation in the next period*. Inflation has been historically between -5% and 15%. When the experiment starts, you have to predict the inflation for the first two periods, i.e. π_1^e and π_2^e. The superscript e indicates that these are predictions. When all participants have made their predictions for the first two periods, the actual inflation (π_1), the output gap (y_1) and the interest rate (i_1) for period 1 are announced. Then period 2 of the experiment begins.

In period 2 you make an inflation prediction for period 3 (π_3^e). When all participants have made their predictions for period 3, the inflation (π_2), the output gap (y_2) and the interest rate (i_2) for period 2 are announced. This process repeats for 50 periods. Therefore, when at a certain period t you make a prediction of the inflation in period $t + 1$ (π_{t+1}^e), the following information is available:

- Values of the actual *inflation, output gap and interest rate* up to and including period $t - 1$;
- Your *predictions* up to and including period t;
- Your *prediction scores* up to and including period $t - 1$.

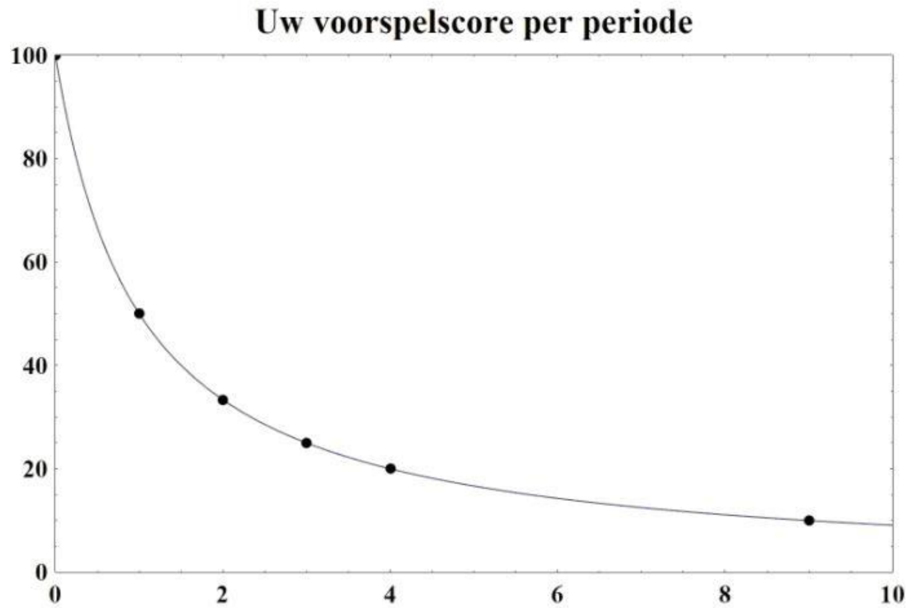
Information about your reward (part 1 of 2)

Your *reward after the experiment has ended* increases with the accuracy of your predictions. Your accuracy is measured by the *absolute error* between your inflation predictions and the true inflation. For each period this absolute error is calculated as soon as the true value of inflation is known; you subsequently get a *prediction score* that decreases as the absolute error increases. The *table below* gives the relation between the absolute predictions error and the prediction score. If at a certain period you predict for example an inflation of 2%, and the true inflation turns out to be 3%, then you make an absolute error of $3\% - 2\% = 1\%$. Therefore you get a prediction score of 50. If you predict an inflation of 1%, and the realized inflation turns out to be -2%, you make a prediction error of $1\% - (-2\%) = 3\%$. Then you get a prediction score of 25. For a *perfect prediction*, with a prediction error of zero, you get a prediction score of 100.

Absolute prediction error	0	1	2	3	4	9
Score	100	50	33/3	25	20	10

Information about your reward (part 2 of 2)

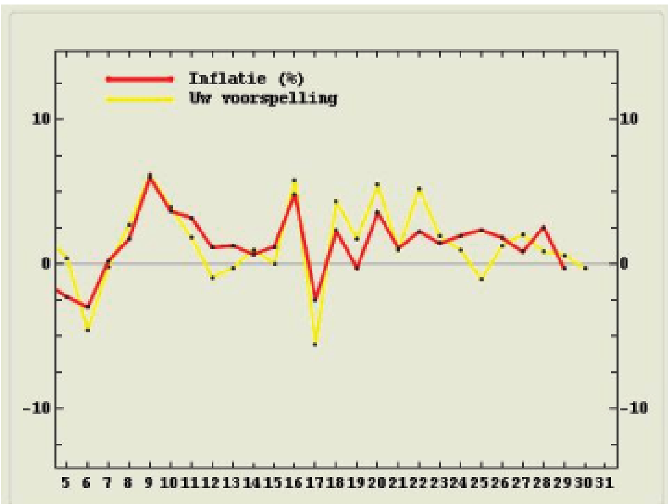
The figure below shows the relation between your *prediction score* (vertical axis) and your *prediction error* (horizontal axis). Notice that your prediction score *decreases more slowly* as your prediction error increases. Points in the graph correspond to the prediction scores in the previous table.



Your *total score* at the end of the experiment consists simply of the sum of all prediction scores you got during the experiment. During the experiment, your scores are shown on your computer screen. When the experiment has ended, you are shown an overview of your prediction scores, followed by the resulting total score. Your *final reward* consists of 0.75 *euro-cent* for each point in your total score (200 points therefore equals 1.50 euro). Additionally, you will receive a *show up fee* of 5 euro.

Information about the computer program (part 1 of 3)

Below you see an example of the *left upper part* of the computer screen during the experiment. It consists of a *graphical representation* of the inflation (red series) and your predictions of it (yellow series). On the horizontal axis are the *time periods*; the vertical axis is in percentages. In the *imaginary situation* depicted in the graph, the experiment is in period 30 and you predict the inflation in period 31 (the experiment lasts for 50 periods). Notice that the graph only shows results of *at most the last 25 periods* and that the *next period* is always on the right hand side.



The *left bottom part* of the computer screen also contains a graph. In this graph the *output gap* and the *interest rate* are shown in the same way as in the above graph.

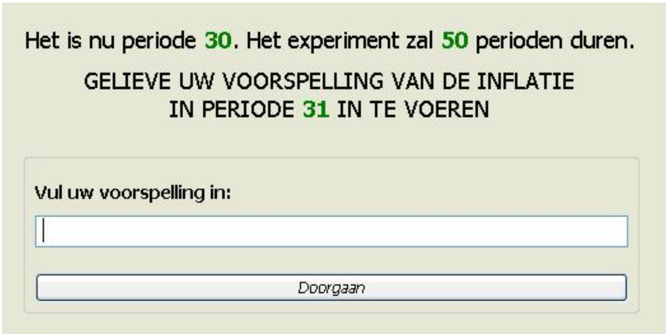
Information about the computer program (part 2 of 3)

Below you see an example of *the right upper part* of the computer screen during the experiment. It consists of a *table* containing information about the results of the experiment in at most the last 25 periods. This information is supplemental to the graphs in the left part of the screen. The *first column* of the table shows the time period (the next period, 31 in the example, is always at the top). The *second and third columns* respectively show the inflation and your predictions of it. The *fourth column* gives the output gap and the *fifth column* the interest rate. Finally, the *sixth column* gives your prediction score for each period separately. Notice that you can use the sheet of paper on your desk to save data longer than 25 periods.

Tijds-periode	Inflatie (%)	Uw Voor-spelling	Output gap (%)	Rente (%)	Voorspel-score
31		???			
30		-0.32			
29	-0.32	0.63	2.15	1.42	56
28	2.54	0.88	2.09	2.13	44
27	0.83	2.04	-7.31	1.71	75
26	1.81	1.28	1.08	1.95	60
25	2.34	-1.09	-2.12	2.09	24
24	1.95	0.99	-1.33	1.99	50
23	1.41	1.9	4.1	1.85	96
22	2.24	5.21	1.39	2.06	24
21	1.07	0.96	-3.92	1.77	55
20	3.56	5.44	5.07	2.39	25
19	-0.3	1.7	-1.89	1.42	78
18	2.29	4.33	1.31	2.07	31
17	-2.49	-5.55	-0.85	0.88	13
16	4.83	5.82	-0.76	2.71	24
15	1.2	0.04	-0.94	1.8	36
14	0.66	0.99	-1.24	1.67	60
13	1.27	-0.29	-0.43	1.82	32
12	1.13	-0.91	-0.18	1.78	27
11	3.21	1.84	3.96	2.3	68
10	3.65	3.91	-1.24	2.41	40
9	5.97	6.13	-4.45	2.99	24
8	1.74	2.71	1.58	1.93	56
7	0.23	-0.14	-3.86	1.56	37
6	-2.98	-4.58	-0.07	0.75	16
5	-2.28	0.43	2.37	0.93	67

Information about the computer program (part 3 of 3)

Below you see an example of the *bottom part* of the computer screen during the experiment. In each period you are asked to submit your inflation prediction in the next period (below *Submit you prediction*). When submitting your prediction, use the *decimal point* if necessary. For example, if you want to submit a prediction of 2.5%, type "2.5"; for a prediction of -1.75%, type "-1.75". Notice that your predictions and the true inflation in the experiment are rounded to two decimals. Moreover, prediction scores are rounded to integers.



(Translation of Dutch)Instructions for participants (output gap forecasters)

Set-up of the experiment

You are participating in an experiment on economic decision-making. You will be rewarded based on the decisions you make during the experiment. The experiment will be preceded by several pages of instructions that will explain how it works. When the experiment has ended, you will be asked to answer some questions about how it went.

- The whole experiment, including the instructions and the questionnaire, is computerized. Therefore you do not have to submit the paper on your desk. Instead, you can use it to make notes.

- There is a calculator on your desk. If necessary, you can use it during the experiment.

- If you have any question during the experiment, please raise your hand, then someone will come to assist you.

General information about the experiment

In the experiment, *statistical research bureaus* make predictions about the *inflation* and the so-called *"output gap"* in the economy. A limited amount of research bureaus is active in the economy. You are a research bureau that makes predictions about *inflation*. This experiment consists of *50 periods* in total. In each period you will be asked to predict the inflation; your *reward after the experiment has ended* is based on the accuracy of your predictions.

In the following instructions you will get more information about the *economy* you are in, about the way in which *making predictions* works during the experiment, and about the way in which your *reward* is calculated. Also, the *computer program* used during the experiment will be explained.

Information about the economy (part 1 of 2)

The economy you are participating in is described by *three variables*: the inflation π_t, the output gap y_t and the interest rate i_t. The subscript t indicates the *period* the experiment is in. In total there are *50 periods*, so t increases during the experiment from 1 through 50.

The *inflation* measures the percentage change in the price level of the economy. In each period, inflation depends on the *inflation predictions* and *output gap predictions* of the statistical research bureaus, and on *minor price shocks*. There is a *positive relation* between the actual inflation and both the inflation predictions and output gap predictions of the research bureaus. This means for example that if the inflation prediction of a research bureaus increases, then actual inflation will also increase (assuming that the other predictions and the price shock remain equal). The minor price shocks have an equal chance of influencing inflation positively or negatively.

Information about the economy (part 2 of 2)

The *output gap* measures the percentage difference between the *Gross Domestic Product (GDP)* and the *natural GDP*. The *GDP* is the value of all goods produced during a period in the economy. The *natural GDP* is the value the total production would have if prices in the economy would be fully flexible. If the output gap is positive (negative), the economy therefore produces more (less) than the natural GDP. In each period the output gap depends on the *inflation predictions* and *output gap predictions* of the statistical bureaus, on the *interest rate* and on *minor economic shocks*. There is a *positive relation* between the output gap and the inflation predictions and output gap predictions, and a *negative relation* between the output gap and the interest rate. The minor economic shocks have an equal chance of influencing the output gap positively or negatively.

The *interest rate* measures the price of borrowing money and is determined by the *central bank*. There is a *positive relation* between the interest rate and the inflation.

Information about making predictions

Your task, in each period of the experiment, consists in predicting the *output gap in the next period*. Inflation has been historically between -5% and 5%. When the experiment starts, you have to predict the output gap for the first two periods, i.e. y_1^e and y_2^e. The superscript e indicates that these are predictions. When all participants have made their predictions for the first two periods, the actual inflation (π_1), the output gap (y_1) and the interest rate (i_1) for period 1 are announced. Then period 2 of the experiment begins.

In period 2 you make an output gap prediction for period 3 (y_3^e). When all participants have made their predictions for period 3, the inflation (π_2), the output gap (y_2) and the interest rate (i_2) for period 2 are announced. This process repeats for 50 periods. Therefore, when at a certain period t you make a prediction of the inflation in period $t+1$ (y^e_{t+1}), the following information is available:

- Values of the actual *inflation, output gap and interest rate* up to and including period $t-1$;
- Your *predictions* up to and including period t;
- Your *prediction scores* up to and including period $t-1$.

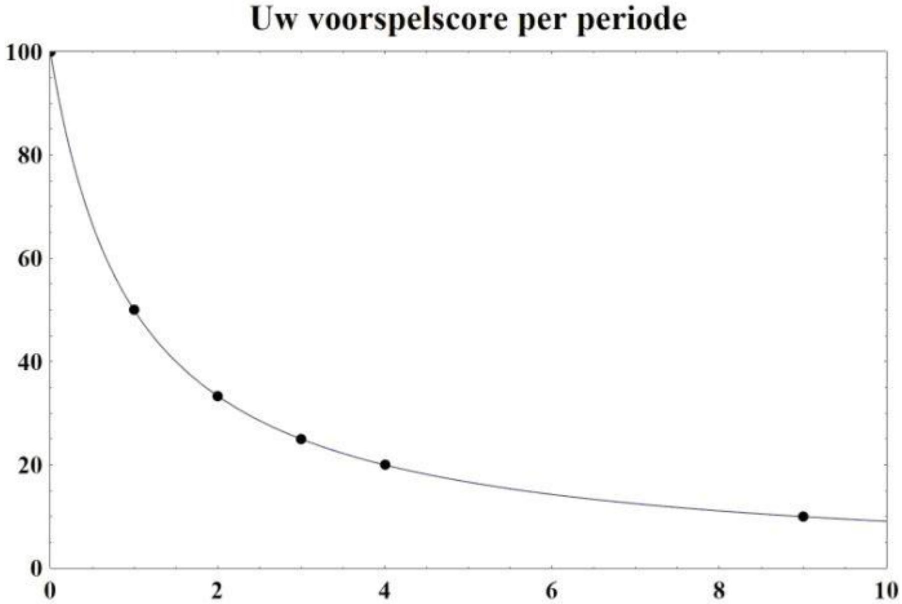
Information about your reward (part 1 of 2)

Your *reward after the experiment has ended* increases with the accuracy of your predictions. Your accuracy is measured by the *absolute error* between your inflation predictions and the true inflation. For each period this absolute error is calculated as soon as the true value of inflation is known; you subsequently get a *prediction score* that decreases as the absolute error increases. The *table below* gives the relation between the absolute predictions error and the prediction score. If at a certain period you predict for example an inflation of 2%, and the true inflation turns out to be 3%, then you make an absolute error of 3% − 2% = 1%. Therefore you get a prediction score of 50. If you predict an inflation of 1%, and the realized inflation turns out to be −2%, you make a prediction error of 1% − (−2%) = 3%. Then you get a prediction score of 25. For a *perfect prediction*, with a prediction error of zero, you get a prediction score of 100.

Absolute prediction error	0	1	2	3	4	9
Score	100	50	33/3	25	20	10

Information about your reward (part 2 of 2)

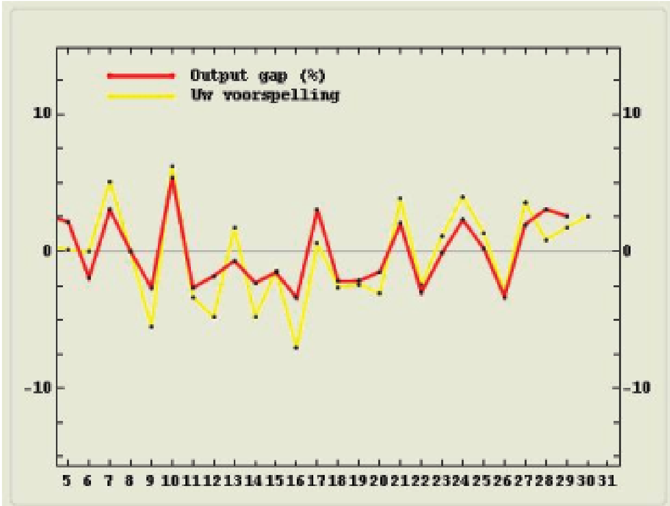
The figure below shows the relation between your *prediction score* (vertical axis) and your *prediction error* (horizontal axis). Notice that your prediction score *decreases more slowly* as your prediction error increases. Points in the graph correspond to the prediction scores in the previous table.



Your *total score* at the end of the experiment consists simply of the sum of all prediction scores you got during the experiment. During the experiment, your scores are shown on your computer screen. When the experiment has ended, you are shown an overview of your prediction scores, followed by the resulting total score. Your *final reward* consists of 0.75 *euro-cent* for each point in your total score (200 points therefore equals 1.50 euro). Additionally, you will receive a *show up fee* of 5 euro.

Information about the computer program (part 1 of 3)

Below you see an example of the *left upper part* of the computer screen during the experiment. It consists of a *graphical representation* of the output gap (red series) and your predictions of it (yellow series). On the horizontal axis are the *time periods*; the vertical axis is in percentages. In the *imaginary situation* depicted in the graph, the experiment is in period 30 and you predict the output gap in period 31 (the experiment lasts for 50 periods). Notice that the graph only shows results of *at most the last 25 periods* and that the *next period* is always on the right hand side.



The *left bottom part* of the computer screen also contains a graph. In this graph the *inflation* and the *interest rate* are shown in the same way as in the above graph.

Information about the computer program (part 2 of 3)

Below you see an example of *the right upper part* of the computer screen during the experiment. It consists of a *table* containing information about the results of the experiment in at most the last 25 periods. This information is supplemental to the graphs in the left part of the screen. The *first column* of the table shows the time period (the next period, 31 in the example, is always at the top). The *second and third columns* respectively show the output gap and your predictions of it. The *fourth column* gives the inflation and the *fifth column* the interest rate. Finally, the *sixth column* gives your prediction score for each period separately. Notice that you can use the sheet of paper on your desk to save data longer than 25 periods.

Information about the computer program (part 3 of 3)

Below you see an example of the *bottom part* of the computer screen during the experiment. In each period you are asked to submit your output gap prediction in the next period (below *Submit you prediction*). When submitting your prediction,

Tijds-periode	Output gap (%)	Uw Voor-spelling	Inflatie (%)	Rente (%)	Voorspel-score
31		???			
30		2.59			
29	2.59	1.78	3.43	2.36	55
28	3.08	0.85	3.44	2.36	31
27	1.98	3.6	2.42	2.11	38
26	-3.3	-2.98	1.38	1.84	76
25	0.24	1.37	4.47	2.62	47
24	2.31	4.02	2.15	2.04	37
23	-0.11	1.15	-0.92	1.27	44
22	-2.99	-2.43	1.	1.75	64
21	2.02	3.87	1.86	1.97	35
20	-1.48	-3.07	0.35	1.59	39
19	-2.14	-2.39	0.48	1.62	80
18	-2.18	-2.59	1.42	1.85	71
17	3.06	0.58	4.01	2.5	29
16	-3.39	-7.06	2.04	2.01	21
15	-1.53	-1.43	4.51	2.63	91
14	-2.28	-4.8	1.44	1.86	28
13	-0.67	1.74	2.78	2.2	29
12	-1.79	-4.8	1.64	1.91	25
11	-2.64	-3.34	1.78	1.95	59
10	5.41	6.25	2.02	2.01	54
9	-2.67	-5.48	1.81	1.95	26
8	0.02	-0.02	0.63	1.66	96
7	3.03	5.15	2.71	2.18	32
6	-1.91	0	-0.7	1.32	34
5	2.17	0.1	-0.76	1.31	32

use the *decimal point* if necessary. For example, if you want to submit a prediction of 2.5%, type "2.5"; for a prediction of -1.75%, type "-1.75". Notice that your predictions and the true inflation in the experiment are rounded to two decimals. Moreover, prediction scores are rounded to integers.

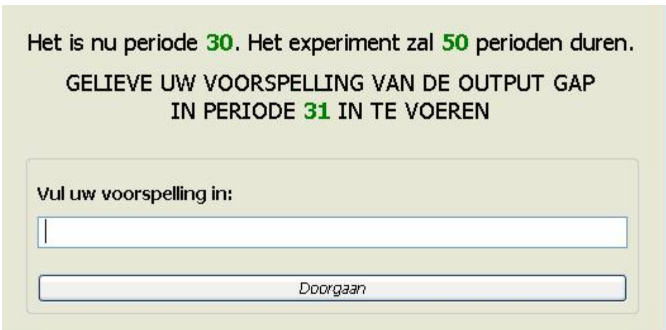


Table 8: MSE over periods $4-50$ of the one-period ahead forecast.

Model	**Tr1*a* gr2**	**Tr1*a* gr3**	**Tr1*b* gr2**	**Tr2*a* gr3**	**Tr2*a* gr2***	**Tr2*a* gr3***	**Tr2*b* gr2**	**Tr3*a* gr1**
RE	*20.421*	0.4557	0.0104	Indeterminacy			0.0339	Indeterminacy
naive	24.334	0.0019	0.0023	6724.1	0.3953	0.1347	0.0058	0.0388
AA	35.418	0.0084	0.0080	7604.8	0.1829	0.1955	0.0089	0.0894
ADA	**24.053**	**0.0015**	***0.0020***	***6115.7***	0.8654	0.2915	0.0283	***0.0320***
WTF	29.856	0.0045	0.0042	8061.6	0.1815	0.0537	**0.0037**	0.0556
STF	71.872	0.0188	0.0167	14721	**0.1285**	**0.0326**	0.0125	0.1345
LAA	35.302	0.0086	0.0077	7364.4	0.1879	0.1829	0.0076	0.0779
4 rules (benchmark)	**39.922**	**0.0019**	**0.0035**	12324	**0.0746**	**0.0202**	***0.0033***	**0.0567**
4 rules (best fit)	29.745	*0.0010*	0.0021	7852.8	*0.0694*	*0.0163*	*0.0033*	0.0401
β	1	3	3	0	10	10	4	1
η	1	1	1	0	0.4	1	0.5	1
δ	0.8	0.6	0.2	0	0.9	0.9	0.9	0.2

Note that the MSE reported for treatment 3 refers to the sum of the MSE relative to inflation and the MSE relative to output gap.
In treatment $2a$ groups 2 and 3 the interest rate hits the zero lower bound respectively in period 43 and period 27, therefore the MSE for Tr2a gr2* and Tr2a gr3* has been computed for periods $4-43$ and periods $4-27$ respectively.
Moreover the MSE for treatment $3a$, group 1, has been computed for periods $4-25$ due to the crash of the session.

Table 9: Out-of-sample performance.

Model	Tr1*a* gr2	Tr1*a* gr3	Tr1*b* gr2	Tr2*a* gr3	Tr2*a* gr2*	Tr2*a* gr3*	Tr2*b* gr2	Tr3*a* gr1
Best Fit Switching Model								
(β, η, δ)	(1,1,0.8)	(3,1,0.6)	(3,1,0.2)	(0,0,0)	(10,0.7,0.9)	(10,0.3,0.7)	(1,0.9,0.1)	(1,1,0.2)
1 p ahead	0.9794	0.0000	0.0003	188.78	0.0023	0.0014	0.0000	0.0008
2 p ahead	22.501	0.0002	0.0002	10602	0.0749	0.1557	0.0007	0.0227
3 p ahead	93.957	0.0000	0.0000	805.38	0.8689	0.7025	0.0000	0.1008
4 p ahead	115.49	0.0010	0.0003	1394.5	3.5035	1.5682	0.0077	0.0522
5 p ahead	84.702	0.0041	0.0000	6297.1	7.7714	3.1128	0.0250	0.0376
6 p ahead	20.159	0.0037	0.0000	13423	11.831	5.5031	0.0413	0.1475
7 p ahead	62.590	0.0060	0.0011	1670.9	16.452	8.4157	0.0287	0.0657
Benchmark Switching Model								
(β, η, δ)(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)	(0.4,0.7,0.9)
1 p ahead	1.6998	0.0000	0.0002	3458.5	0.0501	0.0103	0.0037	0.0033
2 p ahead	27.414	0.0000	0.0000	38333	0.2625	0.1890	0.0068	0.0076
3 p ahead	97.683	0.0000	0.0000	29199	1.3684	0.7643	0.0078	0.0586
4 p ahead	96.953	0.0007	0.0000	28072	3.7197	1.6162	0.0046	0.0087
5 p ahead	43.742	0.0036	0.0002	35032	5.9959	2.8172	0.0000	0.0104
6 p ahead	0.0151	0.0012	0.0007	40493	6.4590	4.3416	0.0001	0.0402
7 p ahead	4.4663	0.0038	0.0015	152250	6.9172	5.8362	0.0003	0.0400
AR(2) Model								
$(\beta_{\pi,0}, \beta_{\pi,1}, \beta_{\pi,2})$	(1.2,0.5,-0.2)	(0.5,0.5,0.1)	(0.8,0.5,0.2)	(-10,0.6,0.2)	(0.7,1.8,-1.0)	(0.2,1.8,-0.9)	(0.6,1.2,-0.5)	(1.3,0.5,0.0)
$(\beta_{y,0}, \beta_{y,1}, \beta_{y,2})$								(0.1,0.7,0.1)
1 p ahead	4.4916	0.0014	0.0004	2944.4	0.0365	0.0060	0.0006	0.0283
2 p ahead	1.7156	0.0000	0.0016	4386.1	0.0773	0.1019	0.0188	0.0094
3 p ahead	9.6386	0.0063	0.0000	0.5891	0.4143	0.4778	0.0063	0.0030
4 p ahead	21.234	0.0046	0.0003	2822.7	0.9962	1.1540	0.0034	0.0035
5 p ahead	11.729	0.0216	0.0292	7238.3	1.6414	2.0189	0.0000	0.0051
6 p ahead	0.9311	0.0070	0.0124	13014	1.4414	3.4115	0.0039	0.0027
7 p ahead	7.2727	0.0132	0.0099	2247.3	1.7650	5.0012	0.0055	0.0336

Note that the squared prediction errors for treatment 3 refer to the sum of the errors relative to inflation and to output gap.
In treatment 2*a* groups 2 and 3 the interest rate hits the zero lower bound respectively in period 43 and period 27, therefore the switching models and the AR(2) model for Tr2*a* gr2* and Tr2*a*gr3* have been estimated on restricted samples of 36 and 20 periods respectively.
Moreover in treatment 3*a* group 1 the switching models and the AR(2) model have been estimated on a sample of 18 periods due to the crash of the session.