*Tiziana Assenza, Peter Heemeijer, Cars Hommes and Domenico Massaro*

# Individual Expectations and Aggregate Macro Behavior

DNB Working Paper

De Nederlandsche Bank
Eurosysteem

# Individual Expectations and Aggregate Macro Behavior

Tiziana Assenza, Peter Heemeijer, Cars Hommes and Domenica Massaro*

# Individual Expectations and

# Aggregate Macro Behavior

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

Cars Hommes[b]    Domenico Massaro[b]

May 11, 2011

[a] *ITEMQ, Catholic University of Milan,*
[b] *CeNDEF, University of Amsterdam,* [c] *Pension Bureau ABN AMRO and University of Amsterdam*

## 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. We find that a more aggressive monetary policy that sets the interest rate more than point for point in response to inflation stabilizes inflation in our experimental economies. We use a simple model of individual learning, with a performance-based evolutionary selection among heterogeneous forecasting heuristics, to explain coordination of individual expectations and aggregate macro behavior observed in the laboratory experiments. Three aggregate outcomes are observed: convergence to some equilibrium level, persistent oscillatory behavior and oscillatory convergence. A simple heterogeneous expectations switching model fits individual learning as well as aggregate outcomes and outperforms homogeneous expectations benchmarks.

# 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 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 most of its appeal in the light of overwhelming 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. Sargent (1999) and Evans and Honkapohja (2001)). Bullard and Mitra (2002), 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), De Grauwe (2010) and Branch and McGough (2009, 2010).

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., Hommes, Sonnemans, Tuinstra, and van de Velden (2005), Adam (2007), Pfajfar and Zakelj (2010); 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 New Keynesian setup. 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. An important novel feature of our experiment is that, in one of the treatments, the aggregate variables inflation and output gap depend on individual expectations of two groups of individuals forming expectations on two different variables, inflation

and the output gap. In particular, we address the following questions:

- are expectations homogeneous or heterogeneous?

- which forecasting rules do individuals use?

- which theory of expectations and learning fits the aggregate as well as individual experimental data?

- which monetary policy rules can stabilize aggregate outcomes in learning to forecast experiments?

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 analyzes the individual forecasting rules used by the subjects, while Section 4 proposes a heterogeneous expectations model explaining both individual expectations and aggregate outcomes. Finally, Section 5 concludes.

# 2 The learning to forecast experiment

In 2.1 we briefly recall the New Keynesian model and then we give a description of the treatments in the experiment. In 2.2 we give an overview of the experimental design and in 2.3 we summarize 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:

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

$$\pi_t = \lambda y_t + \beta \overline{\pi}_{t+1}^e + u_t \,, \tag{2.2}$$

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

where $y_t$ and $\overline{y}_{t+1}^e$ are respectively the actual and average expected output gap, $i_t$ is the nominal interest rate, $\pi_t$ and $\overline{\pi}_{t+1}^e$ are respectively the actual and average expected inflation rates, $\overline{\pi}$ is the inflation target, $\varphi$, $\lambda$, $\beta$ and $\phi_\pi$ are positive coefficients and $g_t$ and $u_t$ are white noise shocks. The coefficient $\phi_\pi$ measures the response of the nominal interest rate $i_t$ to deviations of the inflation rate $\pi_t$ 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 $\overline{y}_{t+1}^e$ and on the real interest rate $i_t - \overline{\pi}_{t+1}^e$. 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 New Keynesian 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 New Keynesian 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. $\overline{y}_{t+1}^e = (1 - \beta)\overline{\pi}\lambda^{-1}$). Given this setup the NK

framework (2.1)-(2.3) specializes to:

$$y_t = (1-\beta)\overline{\pi}\lambda^{-1} - \varphi(i_t - \overline{\pi}^e_{t+1}) + g_t\,, \tag{2.4}$$

$$\pi_t = \lambda y_t + \beta\overline{\pi}^e_{t+1} + u_t\,, \tag{2.5}$$

$$i_t = \phi_\pi(\pi_t - \overline{\pi}) + \overline{\pi}\,, \tag{2.6}$$

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

$$y_t = (1-\beta)\overline{\pi}\lambda^{-1} + \varphi\overline{\pi}(\phi_\pi - 1) - \varphi\phi_\pi\pi_t + \varphi\overline{\pi}^e_{t+1} + g_t\,, \tag{2.7}$$

$$\pi_t = \lambda y_t + \beta\overline{\pi}^e_{t+1} + u_t\,. \tag{2.8}$$

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

$$\pi_t = A_\pi + \frac{\lambda\varphi + \beta}{1 + \lambda\varphi\phi_\pi}\overline{\pi}^e_{t+1} + \xi_t\,, \tag{2.9}$$

where $A_\pi = \dfrac{(1-\beta)\overline{\pi} + \lambda\varphi\overline{\pi}(\phi_\pi - 1)}{1 + \lambda\varphi\phi_\pi}$ 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 $\frac{\lambda\varphi+\beta}{1+\lambda\varphi\phi_\pi}$ 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:

$$y_t = \varphi\overline{\pi}(\phi_\pi - 1) - \varphi\phi_\pi\pi_t + \varphi\overline{\pi}^e_{t+1} + y_{t-1} + g_t, \qquad (2.10)$$

$$\pi_t = \lambda y_t + \beta\overline{\pi}^e_{t+1} + u_t. \qquad (2.11)$$

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

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

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, so that if fluctuations in inflation will arise in the experiment they must be endogenously driven by expectations.

**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

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

where $\mathbf{A}$, $\mathbf{B}$ and $\Omega$ 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 ($\overline{y}^e_{t+1} = (1 - \beta)\overline{\pi}\lambda^{-1}$) 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 a New Keynesian 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[2] ($\phi_\pi = 1.5$).

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 $\phi_\pi$ is equal to 1, the system in Treatments 2 and 3 exhibits a continuum of equilibria.

| | $\phi_\pi$ | $\pi^e_{t+1}$ | $y^e_{t+1}$ | # groups | average earnings $\pi(y)$ in € |
|---|---|---|---|---|---|
| Treatment $1a$ | 1 | $\overline{\pi}^e_{t+1}$ | $(1-\beta)\overline{\pi}\lambda^{-1}$ | 3 | 31 |
| Treatment $1b$ | 1.5 | $\overline{\pi}^e_{t+1}$ | $(1-\beta)\overline{\pi}\lambda^{-1}$ | 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 $(\pi_t)$ 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.

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

$$score = \frac{100}{1+f},$$ (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.
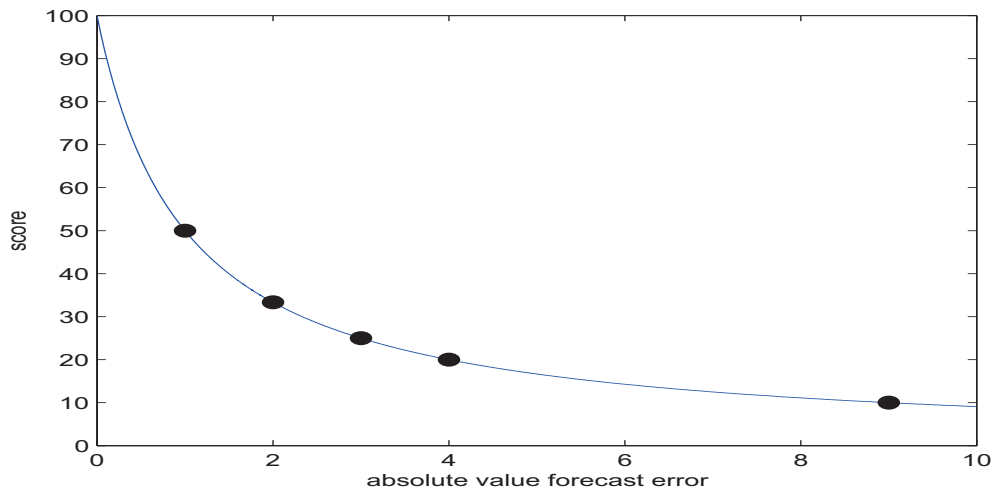


Figure 1: Payoff function

| 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.
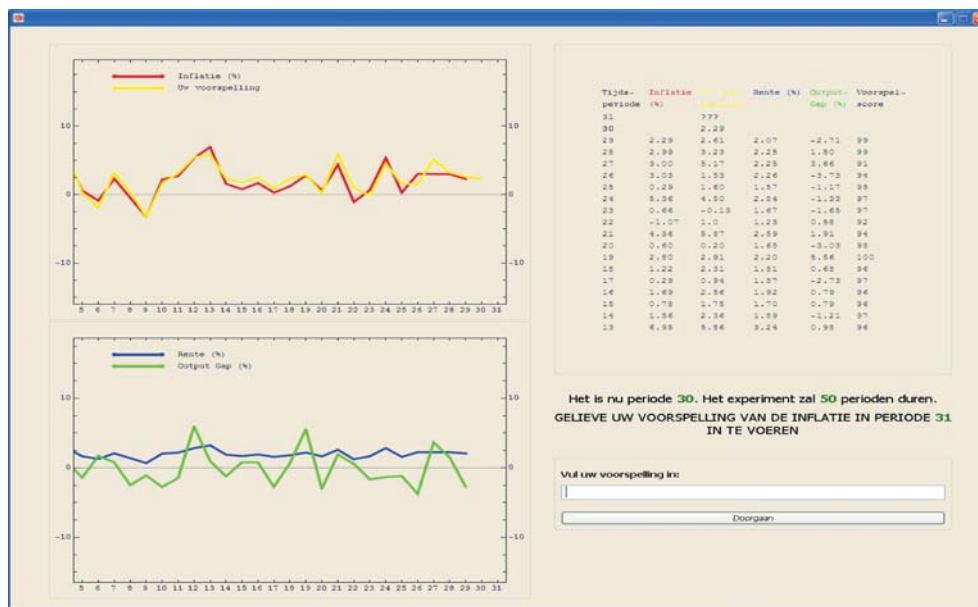


Figure 2: Computer screen for inflation forecasters with time series of inflation forecasts and realizations (top left), output gap and interest rate (bottom left) and table (top right).

## 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. $\beta = 0.99$, $\varphi = 1$, and $\lambda = 0.3$, and we set the inflation target to $\bar{\bar{\pi}} = 2$.

Figure 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].

---

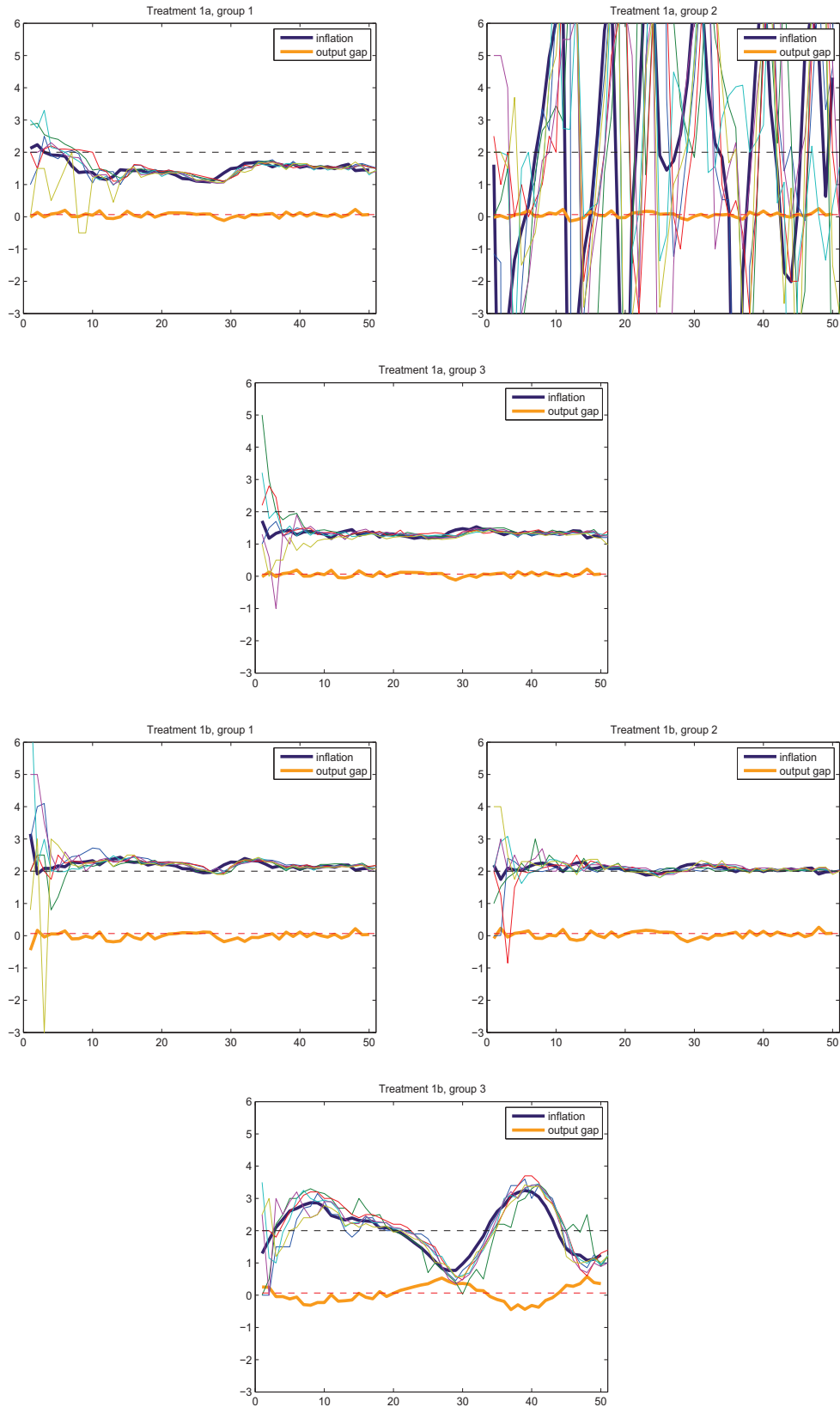[3]The unstable fluctuations were mainly caused by one participant making very high and very low forecasts.

11

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 $2b$ ($\phi_\pi$) 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

---

[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 end effect. In fact, participant 3 predicted an inflation rate of 100% in the last period, causing
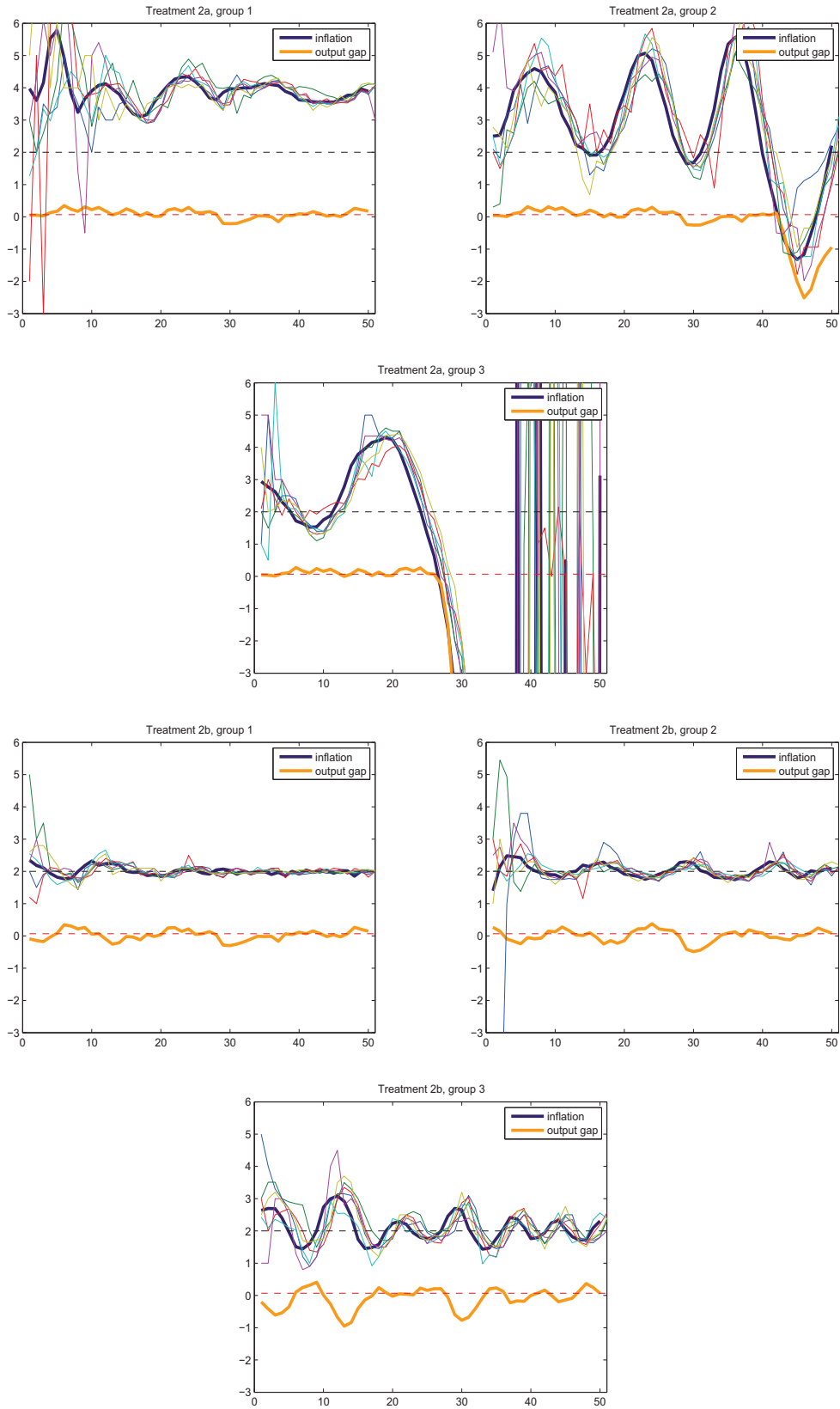
Figure 4: **Upper panels**: Treatment 2a. **Lower panels**: Treatment 2b. Blue thick line: realized inflation; yellow thick line: realized output gap; thin lines: individual forecasts for inflation.
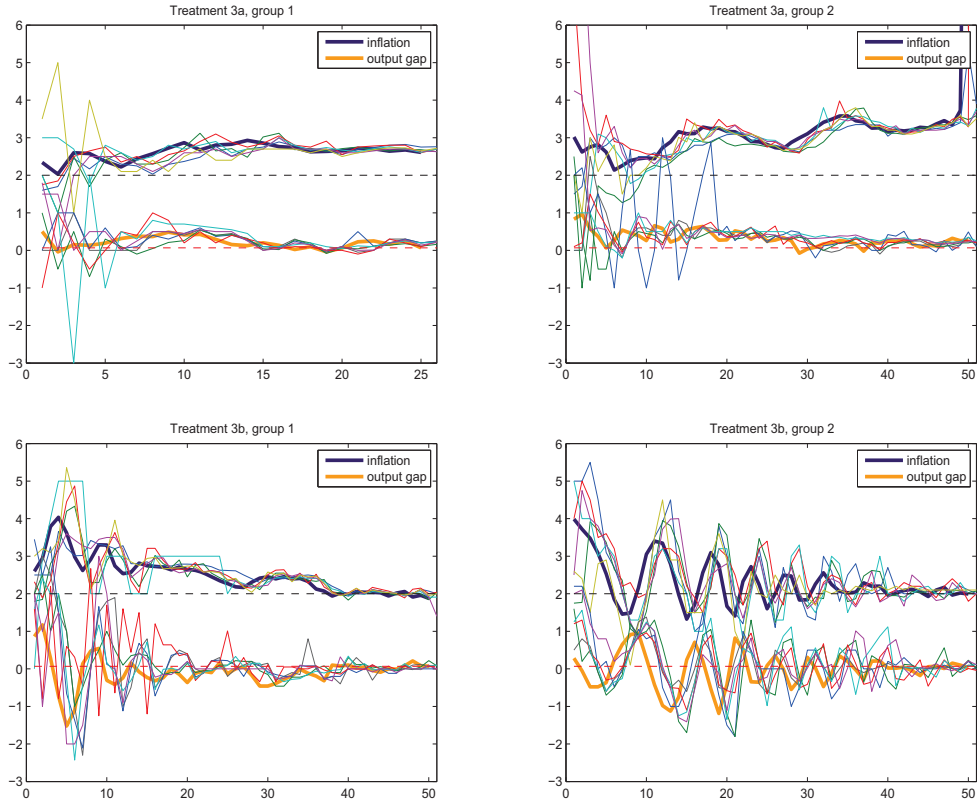
Figure 5: **Upper panels**: Treatment 3a. **Lower panels**: Treateament 3b. Blue thick line: realized inflation; yellow thick line: realized output gap; thin lines: individual forecasts for inflation and output gap.

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

_____

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 Individual forecasting rules

**Estimation of general linear forecasting rules**

For each participant we estimated a simple linear prediction rule of the form

$$\pi_{j,t+1}^e = c + \sum_{i=0}^{2} \alpha_i \pi_{j,t-i}^e + \sum_{i=1}^{3} \beta_i \pi_{t-i} + \sum_{i=1}^{3} \gamma_i y_{t-i} + \mu_t \tag{3.1}$$

$$y_{j,t+1}^e = c + \sum_{i=0}^{2} \delta_i y_{j,t-i}^e + \sum_{i=1}^{3} \varepsilon_i y_{t-i} + \sum_{i=1}^{3} \varsigma_i \pi_{t-i} + \nu_t \tag{3.2}$$

in which $\pi_{j,t+1}^e$ and $y_{j,t+1}^e$ refer to the inflation or output gap forecast of participant $j$ for period $t+1$ (submitted in period $t$). Prediction rule (3.1) applies to inflation forecasters and prediction rule (3.2) to output gap forecasters, using both lagged inflation and lagged output. These prediction rules assume that participants do not use information with a lag of more than three periods; the regression results (see below) show that this is generally a reasonable assumption. We allow for a learning phase, during which participants have not yet fully formed their prediction rules, by leaving the first 11 periods of the experiment out of the regression sample.

Tables 1 – 7 in Supplementary material show the regression results. Of the 102 participants[5] 78% submitted predictions that can be described by a linear rule of the form (3.1) or (3.2) satisfying standard diagnostic tests.[6] In all treatments, the most popular significant regressor is the last available value of the forecasting objective ($\pi_{t-1}$ or $y_{t-1}$). This is followed in most treatments by either the most recent own prediction ($\pi_t^e$ or $y_t^e$) or the second last available forecasting objective ($\pi_{t-2}$ or $y_{t-2}$). Looking at the estimated coefficients, a remarkable property is that all but one of the non-zero coefficients with both the last available forecasting objective and the most recent own prediction are positive suggesting some form of adaptive behavior (see below). In contrast, a clear majority of the non-zero coefficients of remaining

---

[5]The prediction rule specifications (3.1) and (3.2) were applied to all participants except for those in group 3 of Treatment 2a (see Fig. 4) and group 1 of Treatment 3a (see Fig. 5), which experienced respectively total economic collapse and a computer crash during the experiment.

[6]Estimated prediction rules were tested for autocorrelation (Breusch-Godfrey test, 2 lags), heteroskedasticity (White test, no cross terms) and misspecification (Ramsey RESET test, 1 fitted term). A significance level of 5% was used.

regressors, excluding the constant, is negative. Averaging over the participants of all treatments, the number of significant regressors, including the constant, in the estimated prediction rules is 3.4.

The estimation results indicate that most participants, largely irrespective of the treatment they are in, use a consistent, linear prediction rule, at least after a learning phase of 11 periods. What is more, there are clear regularities across groups and treatments regarding the variables the prediction rules are composed of and the sign of their coefficients. Specifically, the fact that the two latest observations of the forecasting objective and the latest own prediction are generally the most used prediction rule components, implies that these variables are of particular importance in the prediction rule specification. The relatively low average number of significant regressors, at 3.4 compared to the 10 potential regressors (including the constant) in (3.1) and (3.2), means that the other variables are used very little as input to form predictions. It may therefore be worthwhile to restrict specifications (3.1) and (3.2) by leaving out these infrequently used regressors. The fact that the estimated non-zero coefficients for the most recent values of the forecasting objective and the own prediction are almost all positive, while the non-zero coefficients of the other variables tend to be negative, similarly suggests that the rule specifications (3.1) and (3.2) are too flexible. Restricting (3.1) and (3.2) along the lines of these regularities could increase the efficiency of the estimates, as well as make the estimated rules easier to interpret from a behavioral viewpoint.

**Estimation of an anchoring-and-adjustment heuristic**

The estimation results of the previous section indicate that the general prediction rules (3.1) and (3.2) can for most participants be restricted without losing much explanatory power. One way of strongly reducing the number of parameters while preserving much of the specifications' flexibility is by fitting an anchoring-and-adjustment heuristic, (Tversky and Kahneman (1974)), named *First-Order*

18

*Heuristic* (FOH)[7]. In the context of our experiment, the FOH has the following form[8]:

$$\pi^e_{h,t+1} = \alpha_1 \pi_{t-1} + \alpha_2 \pi_{h,t} + (1 - \alpha_1 - \alpha_2) \frac{1}{39} \sum_{t=12}^{50} \pi_t + \alpha_3(\pi_{t-1} - \pi_{t-2}) + \mu_t \quad (3.3)$$

$$y^e_{h,t+1} = \gamma_1 y_{t-1} + \gamma_2 y_{h,t} + (1 - \gamma_1 - \gamma_2) \frac{1}{39} \sum_{t=12}^{50} y_t + \gamma_3(y_{t-1} - y_{t-2}) + \nu_t \quad (3.4)$$

The first three terms in (3.3) and (3.4) are a weighted average of the latest realization of the forecasting objective, the latest own prediction and the forecasting objective's sample mean (excluding the learning phase). This weighted average is the (time varying) "anchor" of the prediction, which is a zeroth order extrapolation from the available data at period $t$. The fourth term in (3.3) and (3.4) is a simple linear, i.e. first order, extrapolation from the two most recent realizations of the forecasting objective; this term is the "adjustment" or trend extrapolation part of the heuristic. An advantages of the FOH rule is that it simplifies to well-known rules-of-thumb for different boundary values of the parameter space. For example, the inflation prediction rule (3.3) reduces to Naive Expectations if $\alpha_1 = 1$, $\alpha_2 = \alpha_3 = 0$; it reduces to Adaptive Expectations if $\alpha_1 + \alpha_2 = 1$, $\alpha_3 = 0$. Another special case occurs when $\alpha_1 = \alpha_2 = 0$, so that the anchor reduces to the sample average; we will refer to this case as Fundamentalists, as the sample average is a proxy of the steady state equilibrium level of inflation or output. In the more flexible case $\alpha_1 + \alpha_3 = 1$, $\alpha_2 = 0$ the anchor is time varying; we will refer to this case as a *learning anchor and adjustment* (LAA) rule.

We estimated the FOH rules (3.3) and (3.4) for participants that have a prediction rule of type (3.1) or (3.2) satisfying standard diagnostic tests (see previous Section), and that are not significantly worse described by a FOH rule than by

---

[7]For other applications of the FOH in modeling expectation formation, see Heemeijer, Hommes, Sonnemans, and Tuinstra (2009) and Heemeijer (2009).

[8]In the estimation of (3.3) and (3.4) we included the sample mean of inflation resp. output, which is of course not available to the subjects at the moment of the prediction but acts as a proxy of the equilibrium level. In the heuristic switching model of section 4 we will use the sample average of all previous realizations at every point in time, which generally converges quickly to the sample mean, as an anchor.

a general linear rule. The second criterion was verified by a Wald test on joint parameter restriction. It turns out that 59 out of the 102 participants used in the regression analysis, that is, 58%, pass both criteria. These results are in line with the findings in Adam (2007) where simple forecast functions that condition on a single explanatory variable capture subjects' expectations fairly well. For the participants whose forecasting rules can be successfully restricted to a FOH rule, the estimation results are shown in Tables 8 – 10 in the Supplementary material. Also indicated in Tables 8 – 10 in the Supplementary material are the rules-of-thumb, if any, that the estimated FOH rules are equivalent to (again according to Wald tests). Looking across treatments, the most frequent classifications for the anchors of the prediction rules are Naive and Adaptive Expectations. However, the anchors of almost half of the participants with a FOH rule (26 out of 59) are not equivalent to a well-known rule-of-thumb and are therefore described as "other." Regarding the adjustment part of the estimated FOH rules, it is interesting to see that all participants with a trend extrapolating term in their FOH rules are trend followers, i.e. the coefficient $\alpha_3 > 0$, ranging from 0.3 to 1.4. More than half of the estimated FOH rules (31 out of 59) have a trend-following adjustment term.

Figs. 6(a) – 6(d) illustrate the estimation results in the three-dimensional space $(\alpha_1, \alpha_2, \alpha_3)$. The individual FOH rules are represented by dots in the FOH rule's prismatic parameter space. The prisms show concentrations of dots at the regions corresponding to Naive Expectations ($\alpha_1 = 1$, $\alpha_2 = \alpha_3 = 0$), Adaptive Expectations ($\alpha_1 + \alpha_2 = 1$, $\alpha_3 = 0$), or Trend-Following Expectations ($\alpha_3 > 0$), confirming the classification results above. At the same time, there are substantial differences between the prisms. In particular, Figs. 6(b) and 6(d) show that almost all participants using an FOH rule in Treatments 2a, 2b and 3b have a trend-following adjustment term, while this is much less frequently the case in Treatments 1a, 1b and 3a (see Figs. 6(a) and 6(c)). Comparing with the experimental results (Figs. 3 – 5), the presence of trend-following forecasting rules is clearly related to oscillations in the forecasting objective, which occur more often in Treatments
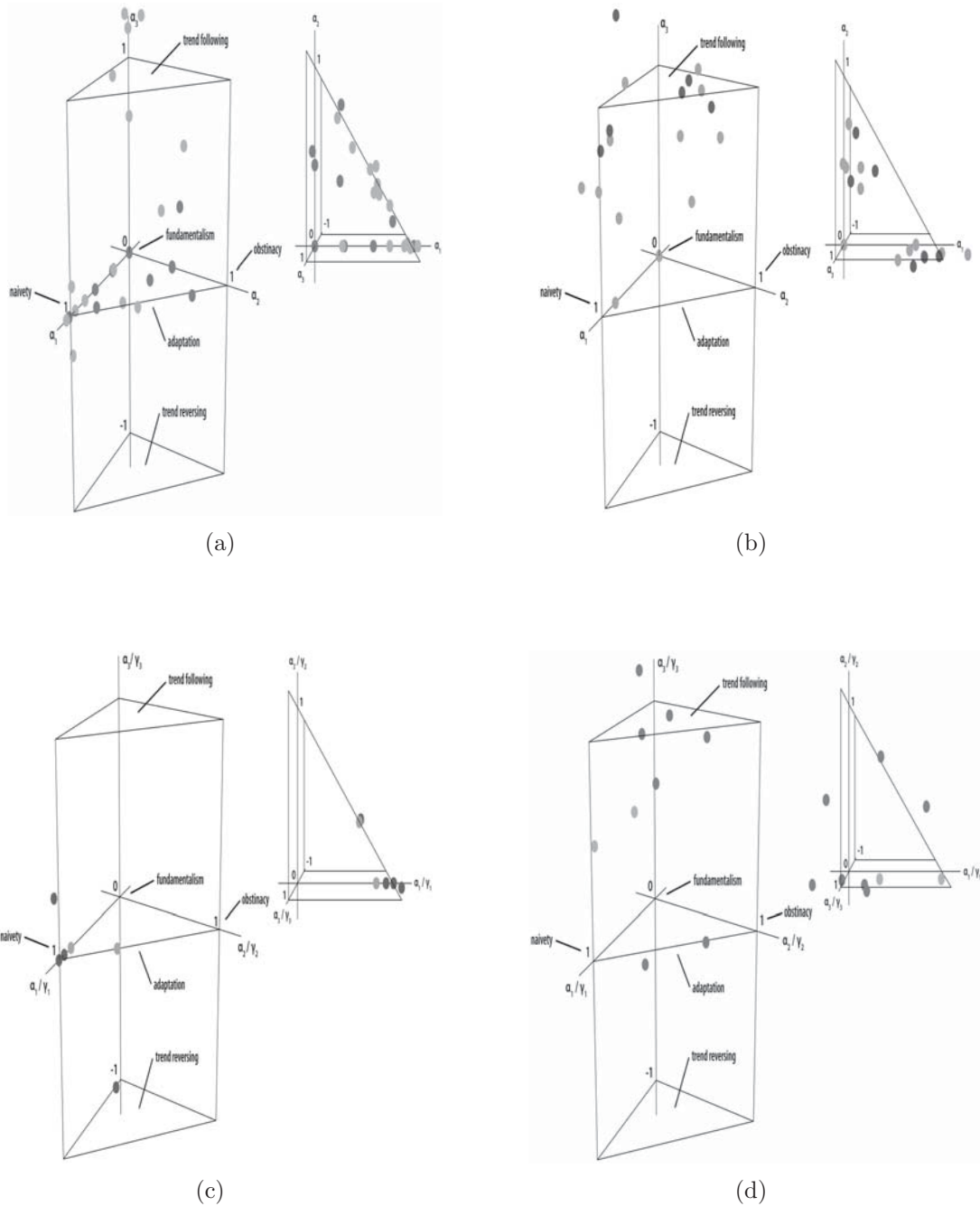
20

Figure 6: Estimated coefficient vectors of First-Order Heuristics (FOH) prediction rules for the participants. The graph on the right of each prism presents a top-down view of the prism. **Top left:** dark dots correspond to participants of Treatment 1a; light dots to participants of Treatment 1b. **Top right:** dark dots correspond to participants of Treatment 2a; light dots to participants of Treatment 2b. **Bottom left:** dark dots correspond to inflation forecasters of Treatment 3a; light dots to output gap forecasters of Treatment 3a. **Bottom right:** Dark dots correspond to inflation forecasters of Treatment 3b; light dots to output gap forecasters of Treatment 3b.

21

2a, 3a and 3b than in the rest of the experiment. Also, an interesting difference between Figs. 6(c) and 6(d) is that Fig. 6(c) contains a cluster of dots close to Naive Expectations, while Fig. 6(d) contains a cluster close to Fundamentalist Expectations (i.e. predictions equal to the forecasting objective's sample mean). Comparing with Fig. 5, it is apparent that the reason for this difference is that inflation and, to a lesser extent, output gap, do not fully converge in Treatment 3a, while they do converge in Treatment 3b. This makes a constant anchor such as in Fundamentalist Expectations more useful in Treatment 3b, and a flexible anchor such as in Naive Expectations more useful in Treatment 3a.

## Graphical evidence for strategy switching

The estimated forecasting rules (3.3) and (3.4) assume a fixed individual prediction rule over the last 40 periods of the experiment. This should be viewed as an approximate forecasting rule; in reality agents may learn and switch to a different forecasting heuristic. This section presents graphical evidence of switching behavior. Fig. 7 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. 7(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. 7(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 switches to a naive forecasting strategy after period 23. In the same experimental session, Fig. 7(c), participant 6 predicting the output gap is using different
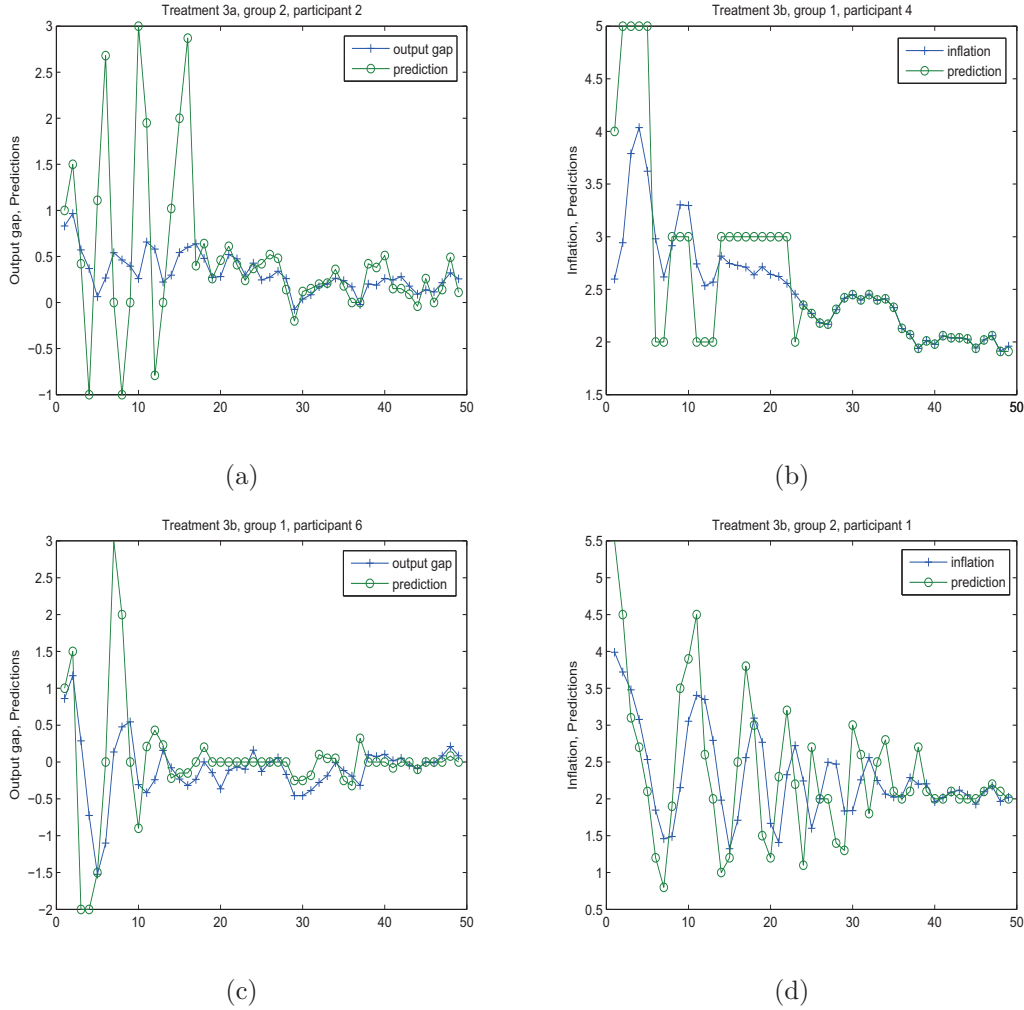
Figure 7: **Individual learning as switching between heuristics.** For every period the subject's forecast $x_{i,t+2}^e$ (green) and the variable being forecast $x_t$, with $x = \pi, y$, are reproduced.

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. 7(d) group 2, treatment 3$b$, 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.

A stylized fact that emerges from the investigation of individual experimental

data is that individual learning has the form of switching from one heuristic to another[9]. Anufriev and Hommes (2009) found a similar result analyzing individual forecasting time series from the asset pricing experiments of Hommes, Sonnemans, Tuinstra, and van de Velden (2005). Moreover, the fact that different types of aggregate behavior, namely convergence to different (non-fundamental) steady states, oscillations and dampening oscillations arise, suggest that heterogeneous expectations play an important role in determining the aggregate outcomes. In the light of the empirical evidence for heterogeneous expectations and individual switching behavior, we introduce in the next section a simple model with evolutionary selection between different forecasting heuristics in order to reproduce individual as well as aggregate experimental data.

# 4   A heterogeneous expectations model

Anufriev and Hommes (2009) 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 $h$.

---

[9]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.

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.

Our goal is to explain three different observed patterns of inflation and output in the experiment: convergence to (some) equilibrium level, permanent oscillations and oscillatory convergence. In order to keep the number of heuristics small, we use a heterogeneous expectation model with only four forecasting rules. These rules, summarized in Table 3, were obtained as heuristics describing typical individual forecasting behavior observed and estimated in our macro experiments. In order to check the robustness of a heterogeneous expectations model across different settings, we fixed the coefficient values to match the set of heuristics used in Anufriev and Hommes (2009) to explain asset pricing experiments. In treatment 3 we apply the same heuristics switching model to both inflation and output forecasting.

Table 3: Set of heuristics

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

## 4.1   50-periods ahead simulations

The model is initialized by two initial values for inflation and output gap, $\pi_1$, $y_1$, $\pi_2$ 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, $\pi_3$ 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: $\beta$, $\eta$, and $\delta$. We used the same set of learning parameters as in Anufriev and Hommes (2009), 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. 8 we used the same realizations for demand and supply shocks as in the experiment and we chose the initial conditions as follows:

- treatment 1*b*, 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 2*b*, 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 3*a*, 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 3*b*, 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. 8 shows realizations of inflation and output gap in the experiment together with the simulated paths using the heuristics switching model[10]. 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[11]. 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[12].

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

| | Treatment | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1b | | 2b | | 3a ($\pi$) | | 3a ($y$) | | 3b ($\pi$) | | 3b ($y$) | |
| | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ |
| 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).

---

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

[11]We reported only simulations for some representative experimental economies which account for the three different aggregate behaviors observed in the experiment. Results for experimental economies with analogous qualitative behavior are similar.

[12]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.
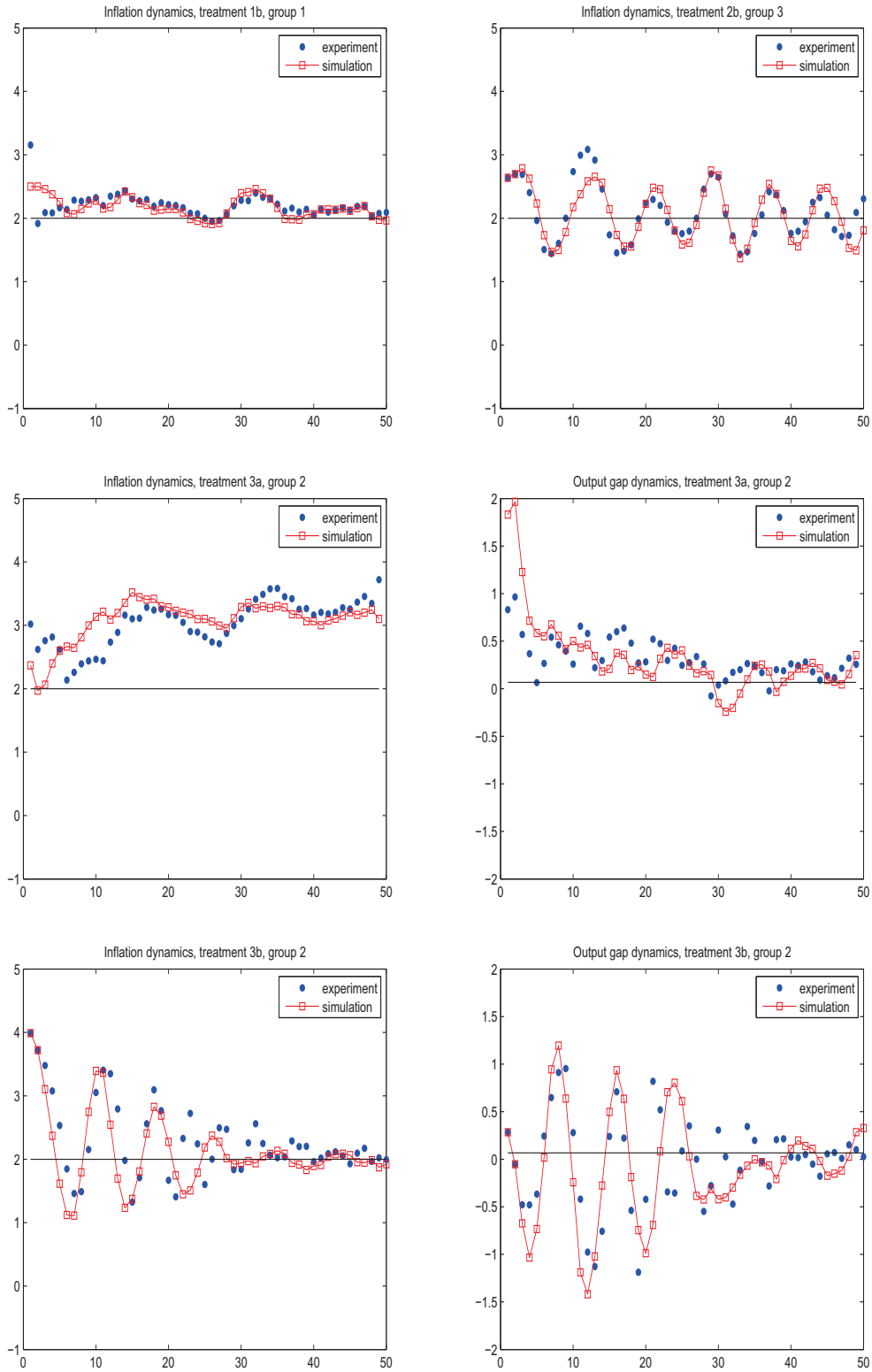
Figure 8: Experimental data (blue points) and **50-periods ahead** heuristics switching model simulations (red lines)

## 4.2 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. 9 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. 10 shows how in different groups different heuristics are taking the lead after starting from a uniform distribution. In treatment 1$b$ group 1 (Fig. 10(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 2$b$ (Fig. 10(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
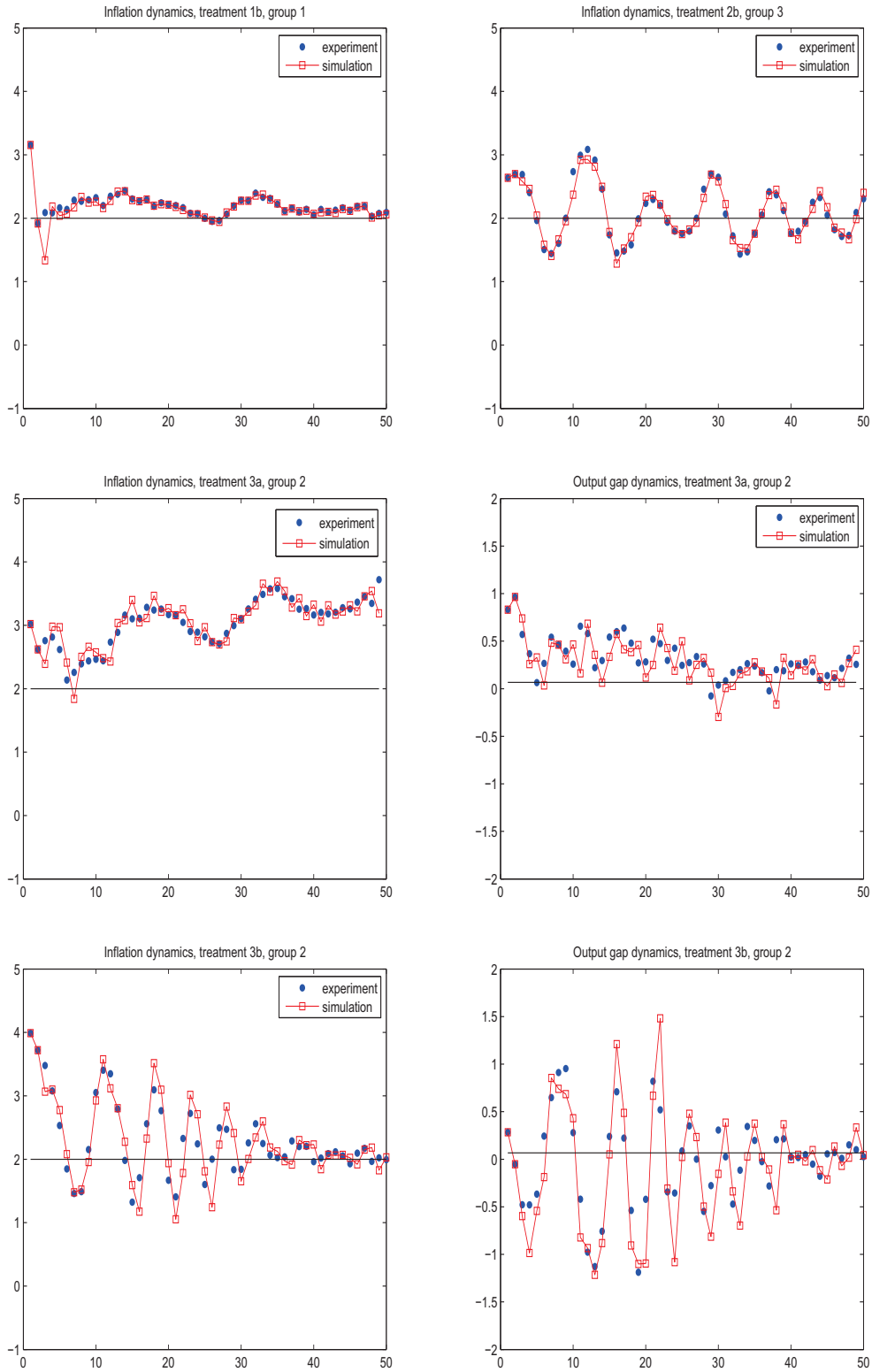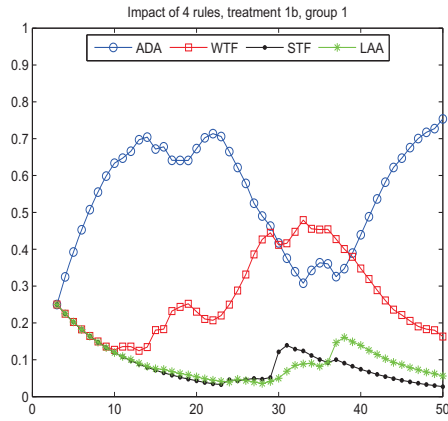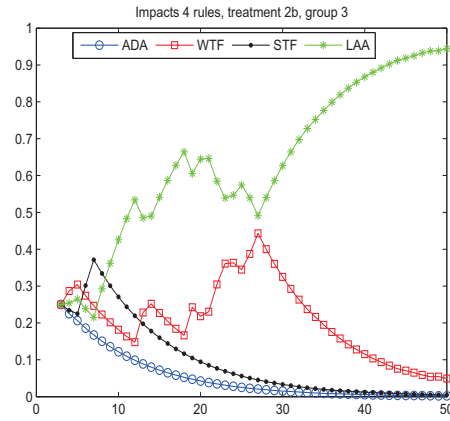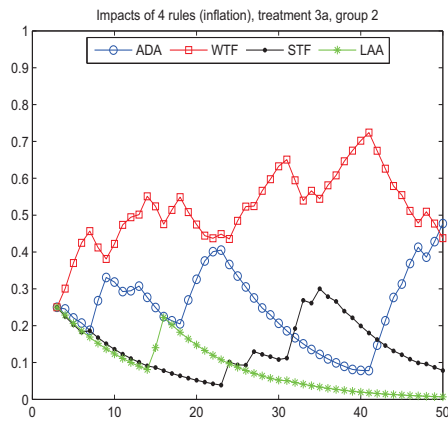
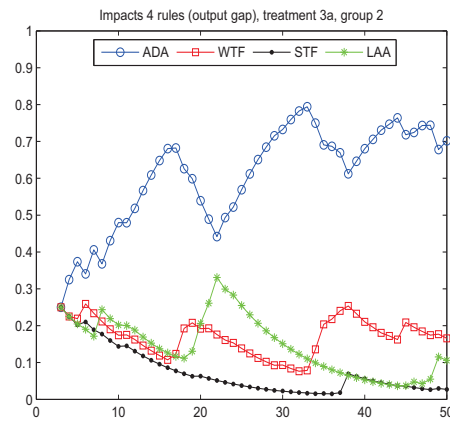Figure 9: Experimental data (blue points) and **one-period ahead** heuristics switching model simulations (red lines)
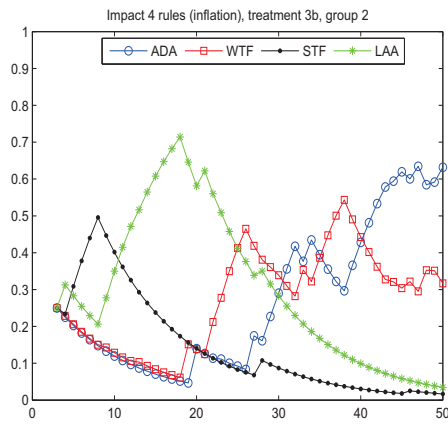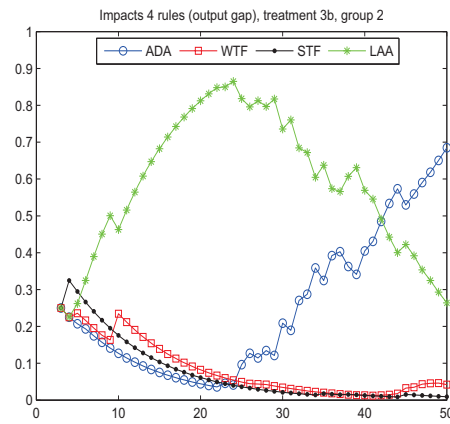
Figure 10: Evolution of fractions of 4 heuristics corresponding to one-period ahead simulations in Fig. 9: adaptive expectations (ADA, blue), weak trend follower (WTF, red), strong trend follower (STF, black), anchoring and adjustment heuristics (LAA, green).

31

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 (2009), our simulation explains oscillatory behavior by coordination on the LAA rule by most subjects. In the early stage of treatment 3*a*, group 2 (Fig. 10(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 (10(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 3*b* group 2 (Fig. 10(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. 10(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. 11 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[13].

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

| | Treatment | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1b | | 2b | | 3a ($\pi$) | | 3a ($y$) | | 3b ($\pi$) | | 3b ($y$) | |
| | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ |
| 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 esti-
mators (Newey-West) have been used to compute standard errors).

---

[13]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
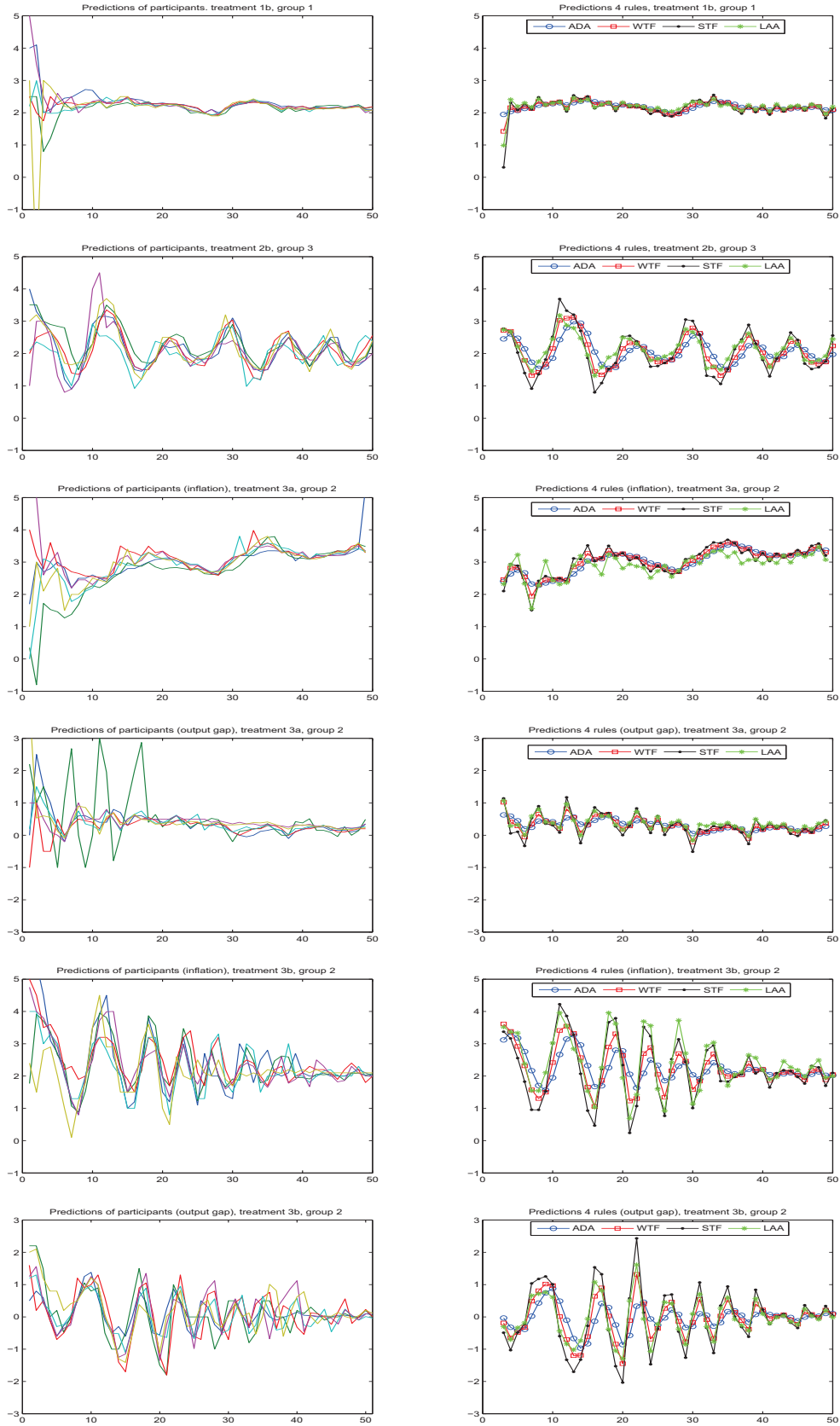of initial conditions.

Figure 11: **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[14] for 9 different models: the rational expectation prediction (RE), six homogeneous expectations models (naive expectations, fixed anchor and adjustment (AA) rule[15], 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[16]. 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[17] cases, 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.

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

[15]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.

[16]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.

[17]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 | Tr1a gr1 | Tr1b gr1 | Tr1b gr3 | Tr2a gr1 | Tr2a gr2 | Tr2b gr1 | Tr2b gr3 | Tr3a gr2 | Tr3b gr1 | Tr3b 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 |
| $\beta$ | 1 | 10 | 2 | 0 | 10 | 10 | 1 | 10 | 7 | 1 |
| $\eta$ | 0.9 | 0.7 | 0.7 | 0 | 0.4 | 0.1 | 0.7 | 0.9 | 0.1 | 0.5 |
| $\delta$ | 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 $\beta$ 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.

### Best Fit Switching Model

| | Tr1a gr1 | Tr1b gr1 | Tr1b gr3 | Tr2a gr1 | Tr2a gr2 | Tr2b gr1 | Tr2b gr3 | Tr3a gr2 | Tr3b gr1 | Tr3b gr2 |
|---|---|---|---|---|---|---|---|---|---|---|
| $(\beta,\eta,\delta)$ | (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

| | Tr1a gr1 | Tr1b gr1 | Tr1b gr3 | Tr2a gr1 | Tr2a gr2 | Tr2b gr1 | Tr2b gr3 | Tr3a gr2 | Tr3b gr1 | Tr3b gr2 |
|---|---|---|---|---|---|---|---|---|---|---|
| $(\beta,\eta,\delta)$ | (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

| | Tr1a gr1 | Tr1b gr1 | Tr1b gr3 | Tr2a gr1 | Tr2a gr2 | Tr2b gr1 | Tr2b gr3 | Tr3a gr2 | Tr3b gr1 | Tr3b gr2 |
|---|---|---|---|---|---|---|---|---|---|---|
| $(\beta_{\pi,0},\beta_{\pi,1},\beta_{\pi,2})$ | (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.

# 5  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. A novel feature of our experimental design is that realizations of aggregate variables depend on individual forecasts of two different variables, 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 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 an economy with heterogeneous individual expectations on two different variables. Simulations of the heuristics switching model show that the model is able to match individual forecasting behavior and nicely reproduce the different observed patterns of aggregate 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 (see Figs. 10(c) and 10(d) where a weak trend following rule dominates inflation forecasting while adaptive expectations dominate output forecasting). 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 target, at least in the long run. In the short run, however, oscillations in inflation and output may arise due to coordination of

individual expectations on trend following rules.

# References

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

ANUFRIEV, M., AND C. H. HOMMES (2009): "Evolutionary Selection of Individual Expectations and Aggregate Outcomes," CeNDEF Working Paper 2009-09, University of Amsterdam.

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 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.

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. 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.

HEEMEIJER, P., C. H. HOMMES, J. SONNEMANS, AND J. TUINSTRA (2009): "Price Stability and Volatility in Markets with Positive and Negative Expectations Feedbacks: An Experimental Investigation," *Journal of Economic Dynamics and Control*, 33, 1052–1072.

HOMMES, C. H. (2011): "The Heterogeneous Expectations Hypothesis: Some Evidence from the Lab," *Journal of Economic Dynamics and Control*, 35, 1–24.

HOMMES, C. H., 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.

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.

TVERSKY, A., AND D. KAHNEMAN (1974): "Judgement under uncertainty: Heuristics and biases," *Science*, 185, 1124–1130.

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

## Previous DNB Working Papers in 2011