# Inflation Expectations and Monetary Policy Design: Evidence from the Laboratory*

Damjan Pfajfar[†]
CentER, EBC, University of Tilburg

Blaž Žakelj[‡]
European University Institute
and Universitat Pompeu Fabra

November 5, 2010

**Abstract**

Using laboratory experiments within a New Keynesian macro framework, we explore the formation of inflation expectations. We find that about 40% of subjects are rational, 35% extrapolate trend, 20% employ adaptive learning and sticky information type models, and 5% behave adaptively. However, rather than using a single model they tend to switch between alternative models. We also study how to design monetary policy in the heterogeneous expectations environment by applying different instrumental rules across treatments. Rules that use actual rather than forecasted inflation produce lower inflation variability and alleviate expectational cycles.

**JEL:** C91, C92, E37, E52

**Key words:** Laboratory Experiments, Inflation Expectations, New Keynesian Model, Adaptive Learning.

*We would like to thank Klaus Adam, Steffan Ball, Bruno Biais, John Duffy, Chryssi Giannitsarou, Cars Hommes, Seppo Honkapohja, Tobias Klein, Aniol Llorente Saguer, Rosemarie Nagel, Charles Noussair, Jan Potters, Juan Manuel Puerta, Emiliano Santoro, and other participants at University of Amsterdam, Universitat Pompeu Fabra, Erasmus University Rotterdam, University of Hamburg, Catholic University in Milan, Tilburg University, 2008 Computational Economics Conference in Paris, 2008 ESA meetings in Lyon, 2008 Nordic Conference on Behavioral and Experimental Economics in Copenhagen and 2008 Learning Week at St. Louis FED for their comments and suggestions. All remaining errors are ours. Previous versions of the paper were circulated as "Experimental Evidence on Inflation Expectation Formation."

[†]Department of Economics, Faculty of Economics and Business Administration, P.O. Box 90153, NL-5000 LE Tilburg, Netherlands. *E-mail*: D.Pfajfar@uvt.nl. *Web*: http://center.uvt.nl/staff/pfajfar/.

[‡]Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain. *Email*: blaz.zakelj@upf.edu. *Web*: http://www.econ.upf.edu/eng/graduates/gpem/zakelj.htm.

# 1 Introduction

This paper discusses an experimental study on the expectations formation process within a macroeconomic framework. Recently, with the development of explicit microfounded models expectations have become pivotal in the modern macroeconomic theory. Central banks increasingly attribute more importance to the developments of households' inflation expectations as they signal future inflationary risks. In line with this development, several theoretical models concerning expectations formation process have been proposed. They postulate informational frictions and heterogeneity of expectations as the main features of the expectation formation process. However, so far these models and their main assumptions have not been subject to rigorous empirical tests. A thorough assessment must rely on micro-level data and the associated distribution, while empirical contributions so far mostly employ aggregate data.[1] Moreover, to evaluate some new theories of expectation formation, e.g. adaptive learning,[2] we need to assure that agents' current information sets encompass all the information from the previous periods. Controlled laboratory environment avoids these methodological issues that are present in the survey data. In this paper we analyze individual data on inflation expectations gathered from an experimental economy and test them for different theoretical models. Insights into households' expectation formation provide useful guidance to central banks how to anchor inflation expectations. After establishing some stylized facts about that we focus on the relationship between policy actions and the formation of inflation expectations. Better understanding of this relationship has been stressed by the Chairman of the Federal Reserve Bernanke (2007) as crucial for the conduct of monetary policy. Advantage of our experiment lies in the usage of the New Keynesian framework and in possibility to compare the aggregate dynamics of inflation and output gap and the effectiveness of monetary policy with the results from the theoretical analysis. We study this question by employing several simple monetary policy rules in different treatments and examine potential implications of the design of monetary policy for forecasting inflation.

This paper provides substantial evidence in support of heterogeneity in the forecasting process. When analyzing individual responses from students of the Universitat Pompeu Fabra and Tilburg University, we find that agents form expectations in accordance with different theoretical models. In our sample approximately $30-45\%$ of agents are rational and around $30-35\%$ of agents predominately extrapolate trend. In addition, $15-25\%$ of agents

[1] Recently, there have been some studies based on micro survey data, e.g. Branch (2004, 2007) and Pfajfar and Santoro (2010). These studies have confirmed that agents only infrequently update their information sets and that they use different theoretical models to forecast inflation.

[2] Adaptive learning assumes that subject are acting as econometricians when forecasting, i.e. reestimating their model each time new data becomes available. See Evans and Honkapohja (2001).

mostly behave in line with adaptive learning and sticky information type models and about $5-10\%$ of agents forecast in an adaptive manner, updating their forecast with respect to the last observed error. Therefore, contrary to the findings of previous experimental studies, we observe a significant proportion of rational subjects. However, as it is not straightforward to define rational subjects, we explore different definitions in order to establish some robustness of our conclusions. Adaptive learning results are also novel as this paper represents one of the first estimations of the gain parameter. This is especially important because experimental data should be regarded as more reliable than survey data. The average gain of agents that employ adaptive learning models is around 0.045. Furthermore, when we allow agents to switch between different models, we find that adaptive learning models are the most popular models for forecasting inflation: 36.7% of all forecasts in our experiment are made with this class of models.

Rather than sticking to one model, switching between alternative models seems to describe subjects' behavior better. We observe that on average subjects switch every 4 periods. Therefore, this paper provides an empirical support for models that postulate endogenous switching, and assume that it is not always optimal to form beliefs in a rational way (e.g. Brock and Hommes, 1997). It could be optimal for some agents in at least some periods to commit to systematic errors as this might be less costly than using a rational rule. Furthermore, we also show that agents use different models as on average in each period 4.5 different models are used in groups of 9 subjects. This suggests that observed heterogeneity is pervasive.

Only a few experimental studies investigate the expectation formation process. The first experiments were performed in a no-feedback environment (e.g. Schmalensee, 1976) and lately some studies have also incorporated a feedback effect in their framework. However, these tend to analyze the expectation processes in an asset pricing setup. Some tests of the rational expectation hypothesis have been conducted within a simple macroeconomic setup (e.g. Williams, 1987; Marimon, Spear, and Sunder, 1993; Evans, Honkapohja, and Marimon, 2001; Adam, 2007).[3] These studies mainly focus on aggregate expectations formation and tend to reject the rational expectations assumption in favor of adaptive way of forming beliefs. On the contrary, we focus on the analysis of micro data and compare them to survey expectations. Our framework allows us to ask the same agents to provide their forecasts over the whole time span. Some analysis of the micro expectations data is conducted by Marimon and Sunder (1995) and Bernasconi and Kirchkamp (2000) in an overlapping generations

[3] See Duffy (2006) for a survey on experimental macroeconomics. Most studies have been conducted in OLG economies with seignorage. Thus our framework is most closely related to the framework of Adam (2007) who studies the expectation formation process in a monetary sticky price environment.

framework. These authors estimate several different regressions in order to study inflation expectation formation and find that most subjects behave adaptively, although Bernasconi and Kirchkamp (2000) provide evidence that adaptive expectations are not of first order degree as argued in Marimon and Sunder (1995). Arifovic and Sargent (2003)[4] also address the issue of inflation expectations formation and study the adaptive hypothesis on individual responses. They also find support of adaptiveness and some evidence of heterogeneity of forecasts.[5] Similar "learning to forecast" experiments are also conducted within the asset pricing framework characterized, as in our case, by positive feedback (see e.g. Hommes et al., 2005 and Haruvy, Lahav, and Noussair, 2007).[6] These studies conclude that most subjects (90%) use simple rules to forecast prices looking at one, two or, at most, three lags of prices.

The baseline experiment described below is repeated under different monetary policy regimes to assess how alternative conducts of monetary policy influence the expectation formation process and the degree of heterogeneity. Monetary policy is modeled using different Taylor-type rules that are commonly used in the literature. Their effectiveness is then compared in terms of variability of inflation and inflation forecasts. We explore how different monetary policy settings anchor inflation expectations. We find that the variability of inflation is significantly affected by the degree of aggressiveness of monetary policy. Our results also suggest that instrumental rules responding to contemporaneous inflation perform better than rules responding to inflation expectations. Furthermore, the design of monetary policy significantly affects the degree of heterogeneity – especially the proportion of trend extrapolation rules – and thus the stability of the main macroeconomic variables. The proportion of trend extrapolation rules increases in an environment characterized by excessive inflation variability and expectational cycles and then further amplifies the cycles. Thus, it is imperative to design a monetary policy that is robust to different expectation formation mechanisms.

Marimon and Sunder (1995) compare different monetary rules in the overlapping generations (OLG) framework to see their influence on the stability of the inflation expectations. In particular, they focus on the comparison between Friedman's k-percent money rule and the deficit rule where the government is fixing the real deficit and finance it through the seigniorage. They provide some evidence in support of Friedman's rule which could help to coordinate agents beliefs and help to stabilize the economy. Similar analysis is also per-

[4] Arifovic and Sargent (2003) focus on the time inconsistency problem, asserting that in many cases policy makers achieve time-inconsistent optimal inflation rate, although in some treatments the economy moves towards sub-optimal (Nash) time consistent outcomes.

[5] Also Fehr and Tyran (2008) suggest that expectations of individuals are heterogeneous. They study the adjustments of nominal prices after the anticipated monetary shock.

[6] See also Hommes (2007) for a short survey.

formed in Bernasconi and Kirchkamp (2000). They argue that Friedman's money growth rule produces less inflation volatility, but higher average inflation compared to constant real deficit rule.[7]

Closer to our framework is the experiment by Adam (2007). He conducts experiments in a sticky price environment where inflation and output depend on expected inflation and analyzes the resulting cyclical patterns of inflation around its steady state. These cycles exhibit significant persistence and he argues that they closely resemble an restricted perception equilibrium[8] where subjects make forecasts with simple underparametrized rules. In our experiment we also detect cyclical behavior of inflation and output gap in some treatments, however we show that these phenomena are not only associated with underparametrization but also with heterogeneity of expectations, the design of monetary policy and (its influence on) the degree of backward-looking behavior.

This paper is organized as follows: Section 2 describes the model for experimental analysis. Section 3 outlines the experimental design. In Section 4 we focus on the analysis of individual responses while in Section 5 we analyze switching dynamics between different models. Section 6 links the results to the monetary policy design and Section 7 concludes.

# 2 Model

In our experiment we use standard forward-looking sticky price New Keynesian (NK) monetary model with different monetary policy reaction functions. The advantage of the NK model is that it is widely used in policy analysis and allows us to compare our experimental results with those obtained in the theoretical literature. Nevertheless, there are two implicit shortcomings in this approach. First, it requires to forecast two periods ahead. It would definitely be easier for participants to produce one period ahead forecast (sometimes called "nowcasting") as they would observe the realizations immediately after their forecasts are made. This would also enable us to simplify the analysis of individual responses, especially in the case of adaptive learning. However, this is not a major obstacle as it is important that we conduct our experiment in a "standard" framework for the analysis of monetary policy. The second drawback is that the forward-looking NK models assume that agents have to forecast both inflation and output gap. To the date, we are not aware of any experiments where subjects are asked to forecast two variables at the same time. This is a considerably more difficult decision to make as we would depart from a standard macro model if we would

[7]The effects of monetary policy design on expectations were also examined in Hazelett and Kernen (2002) were they search for hyperinflationary paths in the laboratory.

[8]Restricted perception equilibrium is generally more volatile than rational expectation equilibrium (for more details see Evans and Honkapohja, 2001).

only ask participants to forecast inflation. Nevertheless, we decided to do this experiment only with expectations of inflation as we were afraid that both issues mentioned in this paragraph would make the task too difficult for individuals. We leave the fully forward-looking NK model for future work.

The baseline framework in the NK approach is a dynamic stochastic general equilibrium model with money, nominal price rigidities, and rational expectations (RE). Lately some authors have augmented this model for adaptive learning and also for heterogeneous expectations (e.g. Branch and McGough, 2009). The model consists of a forward-looking Phillips curve (PC), an IS curve, and a monetary policy reaction function.[9]

The information set at the time of forecasting consists of macro variables at the time $t-1$, although the forecasts are made in period $t$ for period $t+1$. Mathematically we denote this as $E_t\pi_{t+1}$. Strictly speaking, it should be denoted as $E_t\left(\pi_{t+1}|\mathcal{I}_{t-1}\right)$. In fact, $E_t$ (forecast made at period $t$ with information set $t-1$) might not be restricted to just rational expectations.

The IS curve is specified as follows:

$$y_t = -\varphi\left(i_t - E_t\pi_{t+1}\right) + y_{t-1} + g_t, \tag{1}$$

where interest rate is $i_t$, $\pi_t$ denotes inflation, $y_t$ is output gap, and $g_t$ is an exogenous shock. The parameter $\varphi$ is the intertemporal elasticity of substitution in demand. We can observe that we do not have expectations of output gap in the specification. Instead, we have lagged output gap.[10] Compared to purely forward-looking specifications, our model might display more persistence in output gap. This is the most significant departure from otherwise standard macroeconomic model.

Aggregating across the price setting decisions of individual firms yields the linear relationship in the equation (2). Thus, the supply side of the economy is summarized in the following PC:

$$\pi_t = \beta E_t\pi_{t+1} + \lambda y_t + u_t. \tag{2}$$

The longer prices are fixed on average, i.e. the smaller is $\lambda$, the less sensitive inflation is to the current output gap. The parameter $\beta$ is the subjective discount rate. The shocks $g_t$

[9] Detailed derivations are in, e.g., Woodford (1996), or textbooks such as Walsh (2003) or Woodford (2003).

[10] In principle, one could argue that this specification of IS equation corresponds to the case when subjects have naive expectations on output gap or it is assumed the extreme case of habit persistence. The main reason for including lagged output gap in our specification is that we want another endogenous variable to influence the law of motion for inflation. Furthermore, we prefer that even in the case when agents have rational expectations they have to use the observed information on output gap for forecasting inflation as it enters into the perceived law of motion of the rational expectations form.

and $u_t$ are unobservable to subjects and follow the following process:

$$\begin{bmatrix} g_t \\ u_t \end{bmatrix} = \Omega \begin{bmatrix} g_{t-1} \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} \widetilde{g}_t \\ \widetilde{u}_t \end{bmatrix};$$
$$\Omega = \begin{bmatrix} \kappa & 0 \\ 0 & \nu \end{bmatrix},$$

where $0 < |\kappa| < 1$ and $0 < |\nu| < 1$. $\widetilde{g}_t$ and $\widetilde{u}_t$ are independent white noises, $\widetilde{g}_t \backsim N\left(0, \sigma_g^2\right)$ and $\widetilde{u}_t \backsim N\left(0, \sigma_u^2\right)$. In the NK literature it is standard to assume AR(1) shocks. $g_t$ could justified as a government spending shock or a taste shock and standard interpretation of $u_t$ is the technology shock. All these shocks are found to be quite persistent in the empirical literature (see e.g. Cooley and Prescott, 1995 or Ireland, 2004). In the experimental context it is important to have some exogenous unobservable component in the law of motion for endogenous variables, so that we prevent the extreme case where all agents coordinate on the forecasts identical to inflation target. If we would not have AR(1) shock this would represent the dominant strategy. This is especially relevant concern as we initialize the model in the rational expectations equilibrium (REE).

## 2.1 Monetary Policy Reaction Functions

To close the model, we have to specify the interest rate rule.[11] We use two alternative Taylor-type rules in different treatments. Most of our attention is devoted to forward-looking reaction functions: inflation forecast targeting where interest rate is set in response to inflation expectations. We study three parametrizations of this rule and investigate how different degrees of central bank's aggressiveness in stabilizing inflation influence inflation expectations. Next, we ask whether it is better for the central bank to respond to the current or expected inflation. Therefore, we also analyze the pure inflation targeting.

We start with the following interest rate rule (Inflation Forecast Targeting):

$$i_t = \gamma\left(E_t \pi_{t+1} - \overline{\pi}\right) + \overline{\pi}. \tag{3}$$

In this version the central bank responds to deviations of inflation from the target, $\overline{\pi}$. We vary $\gamma$ in different treatments and study stability of the system under alternative reaction

[11] Engle-Warnick and Turdaliev (2006) investigate the conduct of monetary policy in an experimantal setting. Their subjects are only told to act as policymakers and to stabilize inflation. Most of the subjects control inflation relatively well and authors argue that Taylor rules provide a good description of subjects' policy decisions.

coefficients attached to inflation.

The second alternative specification is inflation targeting, where the monetary authority is assumed to respond to deviations of contemporaneous inflation from the inflation target:

$$i_t = \gamma \left(\pi_t - \overline{\pi}\right) + \overline{\pi}. \tag{4}$$

## 2.2 Calibration

We use McCallum and Nelson (2004) calibration. This calibration represents one of the standard calibrations for the NK models. In order to have inflation in positive numbers for most of the periods we set the inflation target to $\overline{\pi} = 3$. A summary of the calibration is reported in the next table.

Insert Table 1 about here

Treatments are fully comparable as we have exactly the same shocks in all treatments. In particular, $\kappa$ and $\nu$ are calibrated to 0.6, while their standard deviations are 0.08.

# 3 Experiment

## 3.1 Design[12]

Experimental subjects participated in a simulated economy of 9 agents.[13] Each session of a treatment has 2 independent groups ("economies"), therefore 18 subjects participate in each session. All participants were recruited through a recruitment programs for undergraduate students at the Universitat Pompeu Fabra and University of Tilburg. Invitations to apply were sent to all of around 1300 students in a database at Pompeu Fabra and to about 1200 students at Tilburg, except to those that already participated in one of our sessions before. There are 70 periods in each treatment. We scaled the length of each decision sequence and number of repetitions in a way that each session lasts approximately 90 to 100 minutes, including the time for reading the instructions and 5 trial periods at the beginning. The program is written in Z-Tree experimental software (Fischbacher, 2007).

Subjects are presented with a simple fictitious economy setup. As it is shown above, the economy is described with three macroeconomic variables: inflation, output gap and interest

[12]Experimental instructions can be found in Appendix B.

[13]The common view among the experimental economists is that we do not need many subjects in the microfounded experiments. Most of the learning to forecast experiments are conducted with 5-6 subjects, e.g. Hommes, Sonnemans, Tuinstra, and van de Velden (2005), Adam (2007), Fehr and Tyran (2008).

rate. Participants observe time series of these variables in a table, up to the period $t-1$. 10 initial values (periods $-9, \ldots, 0$) are generated by the computer under the assumption of rational expectations. Subjects' task is to provide inflation forecasts for the period $t+1$. The underlying model of the economy is unknown to them. We explain the meaning and relevance of the main macroeconomic variables and inform them that their decisions have an impact on the realized output, inflation and interest rate in time $t$.

In every period $t$, there are two decision variables subjects have to input: *i)* prediction of the $t+1$ period inflation; *ii)* 95% confidence interval of their inflation prediction. In 4 out of 6 independent groups in each treatment subjects have to report the interval as a number of percentage points for which the actual inflation can be higher or lower. In the other 2 groups in each treatment, subjects are simply asked for the lower and the upper bound of their inflation prediction interval.

After each period subjects receive information about the realized inflation in that period, their prediction of it, and the payoff they have gained. Subjects' payoffs depend on the accuracy of their predictions. The accuracy benchmark is the actual inflation rate computed from the underlying model on the basis of predictions made by all agents in the economy. In the subsequent rounds subjects are also informed about their past forecasts. They do not observe the forecasts of other individuals and their performance. The payoff function, $W$, is a sum of two convex components as described below:

$$
\begin{aligned}
W &= W_1 + W_2, \\
W_1 &= \max\left\{\frac{1000}{1+f} - 200,\ 0\right\}, \\
W_2 &= \max\left\{\frac{1000x}{1+CI} - 200,\ 0\right\}, \\
x &= \begin{cases} 1 & \text{if} \quad CI \geq f \\ 0 & \text{if} \quad otherwise \end{cases}, \\
f &= \left|\pi_t - E_{t-1}\pi_t\right|.
\end{aligned}
$$

The first, $W_1$, depends on their forecast errors and is designed to encourage subjects to give accurate predictions. It gives subjects a payoff if their forecast errors, $f$, are smaller than 4. The second, $W_2$, depends on the width of their confidence interval and intends to motivate subjects to think about the variance of actual inflation since it is more rewarding when it is narrower. $CI$ is either equal to their point estimate of confidence interval or half of the difference between upper and lower bound. They receive a reward if their confidence intervals, $CI$, are not larger than $\pm 4$ percentage points, conditional on the fact that actual inflation falls in the given interval: $CI \geq |\pi_t - E_{t-1}\pi_t|$. With this setup we restrict to positive payoffs. Compared to more standard quadratic payoff functions, ours gives greater reward to more accurate predictions. Similar approach is used in Adam (2007).

Participants received detailed instructions before the experiment started. To ensure understanding of the task, we read instructions out loud and present their task descriptively along with examples. We accompanied the payoff function with generous explanation and a payoff matrix on a separate sheet of paper. Subjects also filled in a short questionnaire after they have read the instructions and answered the questions about the procedures to make sure that all participants understood them.

## 3.2 Treatments

The experiment consists of 5 sessions (a pilot session and 4 regular sessions). Participants on average earn around €15, depending on treatment and individual performance. Every experimental session represents a different treatment, each using a different specification of monetary policy reaction function.

Insert Table 2 about here

The first three treatments, as shown in the Table 2, deal with the parametrization of the inflation forecast targeting given in equation (3). In this setup, the coefficient $\gamma$ determines central bank' aggressiveness to deviations of inflation from its target. It is also believed that the higher the $\gamma$ is, the stronger is the stabilizing effect of the monetary policy rule. It is of our key interest to see how subjects react to more and less aggressive interest rate policies. Moreover, we test in a controlled environment whether different slope coefficients indeed have the expected stabilization effect.

Majority of empirical findings agree that the magnitude of the slope coefficient is around 1.5. Generally, when $\gamma > 1$ the interest rate rule is E-stable and produces a determinate outcome[14] (Taylor principle) while the one with $\gamma \leq 1$ is E-unstable and indeterminate. When Taylor principle holds all our treatments yield determinate and E-stable REE. Initially, we planned to perform a treatment with $\gamma < 1$ to check whether this leads to instability, however findings from the pilot treatments convinced us this is not a suitable choice as subjects quickly reached extremely high levels of inflation. This clearly leads to explosive behavior of the system, so our findings suggest that Taylor principle holds.[15]

[14] E-stability is asymptotic stability of an REE under least sqares learning. Under determinacy we mean the existence of a unique dynamically stable REE. For more detailed definition see Evans and Honkapohja (2001). Proof that this is also the case in our setup can be obtained from the authors upon request.

[15] Moreover, under these circumstances inflation never returned to the target inflation and just keept growing. Therefore the effect of output gap on inflation never outweights the expected inflation effect.

For our first and benchmark treatment we decided to follow Taylor and chose $\gamma = 1.5$. Average behavior of groups in the first treatment show no convergence to target inflation, so we choose $\gamma = 1.35$ as sufficiently different case for a comparison. Alternatively, we chose $\gamma = 4$ as parametrization with high stabilizing effect where convergence to the target inflation should be faster.

In treatment 4 we focus on what measure of inflation should central banks target: the expected inflation by subjects or actual inflation. We perform a treatment using inflation targeting rule where central bank reacts to current inflation, with $\gamma = 1.5$ as in our benchmark case.

# 4 Analysis of Individual Inflation Forecasts

The analysis of individual responses focuses in the first part on learning dynamics. Several learning models are simulated in order to find the best fit of each individual series on expectations. We also estimate other standard models of expectation formation including common rationality tests. All these models are estimated with pooled OLS techniques where we estimate individual specific coefficients. Reported results are with robust standard errors that, where appropriate, take into account the presence of clusters in groups (or treatments). Below we present each of these tests and briefly comment estimation for all subjects while in the discussion we determine the best performing model for each subject. In the next section we dig deeper and investigate potential switching of subjects between different models.

In 4 treatments of our experiment and 24 independent groups we gathered $40,320$ data points from 216 subjects. The mean inflation forecast for all treatments is around 3.06% and the mean inflation is 3.02% where the inflation target is set to 3%. Standard deviations of inflation and inflation expectations vary substantially across groups. For inflation expectations the largest is 6.31 and the lowest 0.23 while for inflation the largest is 5.83 and the smallest is 0.24. Standard deviations of inflation forecasts are usually higher than standard deviations of inflation for groups with higher volatility while for groups with lower volatility this might not necessary be the case. Figure A1 in the Appendix A displays distribution of inflation forecasts in each treatment.

Insert Table 3 about here

In Figure 1[16] it is possible to distinguish signs of rounding effect (or digit preference). This is especially evident for the responses bellow 0 and above 6, where we can observe a

[16] The full range of responses reported is between $-13.9$ and 24, however in this histogram we restrict to responses between $-3$ and 10.

clear pattern that resembles rounding: the frequency of responses are significantly higher for round numbers than for the neighboring decimal numbers. A closer inspection reveals that rounding is also present for the responses between 0 and 6, only that rounding here does not take place only for responses such as 2, 3 and 4, but also for 2.5 and 3.5. This is due to the fact that in treatments where variability is lower subjects round on the basis of a smaller grid. Overall, we can point out that 72% of all responses are reported to one decimal point accuracy, while 13% of them are to the accuracy of 2 decimal points. The remaining 15% of forecasts are rounded as integers. The overall share of the latter is significantly higher for the groups with higher volatility compared to the groups displaying lower volatility.

Insert Figure 1 about here

However, we have to point out that survey data usually display more rounding, particularly the Michigan survey (see Curtin, 2005, Bryan and Palmqvist, 2005). Subjects in experiments are paid according to their performance and thus the accuracy of forecasts always matters. On the contrary, in survey data we can observe the effect of inattentiveness[17] when inflation is low and stable. In this environment it can be said that the forecast accuracy is relatively less important than in the periods when inflation is more volatile and higher. The mean of forecast errors in our experiment is 0.04 and the standard deviation is 1.23. Thus, there is only a slight positive bias of errors. Furthermore, subjects overpredict in 51.2% cases and underpredict in 48.8%.

For the treatments where subjects provide a single symmetric confidence bound the range of responses is between 0 and 8.3, although subjects know that inputs larger than 4 do not result in any payoff. The average confidence interval is 0.61 and a standard deviation is 0.69. In the treatments where subjects provide upper and lower confidence bounds, the mean difference between inflation prediction and lower interval bound is 0.37 and standard deviation is 0.31; range of responses is between 0 and 4.6. The upper confidence bound differences range from 0 to 20.2, the mean being 0.41 and standard deviation being 0.54. This suggests that subjects are on average more likely to expect an inflation increase than an inflation decrease. It is also interesting to see how accurate experimental subjects are in determining the confidence bounds. Thaler (2000) finds that "when people asked about their 90% confidence limits, the answers will lie within the limits in less than 70% of the time" (p. 133). Giordani and Söderlind (2003) get very similar result (72%). Our results confirm Thaler's hypothesis that people on average underestimate risk in an even stronger manner. 60.5% of the times subjects managed to set confidence bounds that included the actual inflation in the next period.[18]

[17] Inattentiveness was first discussed by Mankiw and Reis (2002).

[18] Our instructions required subjects to introduce their prediction with 95% confidence bounds which

### 4.1 Tests of Rational Expectations

Several econometric tests are designed to check the rationality of forecasts. In this subsection we apply some standard tests commonly employed in the survey data literature.[19] We assess different degrees of forecast efficiency and check whether forecasts yield predictable errors. The simplest test of efficiency is a test of bias:

$$\pi_{t+1} - \pi^k_{t+1|t} = \alpha, \tag{5}$$

where $\pi_{t+1}$ is inflation at time $t+1$ and $\pi^k_{t+1|t}$ is $k^{th}$ subject's inflation expectations for time $t+1$ made at time $t$ (with information set $t-1$). By regressing expectational errors on a constant we check whether inflation expectations are centred around the right value. Majority of agents produce unbiased estimates of inflation. Overall, only 7.9% of them produce biased estimates at a 5% significance level and only 4.6% at a 1% threshold. Most of them are from treatments 2 and 4.

The next regression represents a further test for rationality:

$$\pi_{t+1} = a + b\pi^k_{t+1|t}. \tag{6}$$

As in Mankiw, Reis, and Wolfers (2004) the last expression can be simply augmented to test whether information in forecasts are fully exploited:

$$\pi_{t+1} - \pi^k_{t+1|t} = a + (b-1)\,\pi^k_{t+1|t}, \tag{7}$$

where rationality implies jointly that $a = 0$ and $b = 1$. As in the test for bias, under the null of rationality these regressions are meant to have no predictive power. The latter model is a more strict test of rationality and is seldomly fulfilled in the survey data literature. On the contrary, our results suggest that 28.7% of agents exploit all the available information at a 5% significance level and 42.1% of them when we decrease the threshold to 1%. Treatment 2 is associated with the highest proportion of rational agents (48% and 57%, accordingly). Compared to other experimental studies, these tests suggest that a significant proportion of subjects behave rationally, although in asset pricing experiments Heemeijer, Hommes, Sonnemans, and Tuinstra (2007) find a significant proportion of fundamental traders. These can be associated with rational expectations. Also Roos and Luhan (2008) show that about

makes the underestimation claim even stronger. More analysis on confidence intervals can be found in our companion paper, Pfajfar and Žakelj (2010).

[19] See Pesaran (1987), Mankiw, Reis, and Wolfers (2004) and Bakhshi and Yates (1998) for a review of these methods.

23% of subjects do not have biased price expectations.[20]

## 4.2 Sticky Information Type Regression

In this section we estimate a simple weighted average regression similar in formulation to sticky information model by Carroll (2003a) and adaptive expectations. In our framework we have forecasts derived under the assumption of rational expectations while Carroll (2003a) implements professional forecasters predictions. We estimate the following equation:

$$\pi^k_{t+1|t} = \lambda_1 \pi^{RE}_{t+1|t} + (1-\lambda_1)\,\pi^k_{t|t-1}; \tag{8}$$

$$\pi^k_{t+1|t} = \lambda_1 \eta_0 + \lambda_1 \eta_1 y_{t-1} + (1-\lambda_1)\,\pi^k_{t|t-1}, \tag{9}$$

where $\pi^{RE}_{t+1|t}$ is a rational expectations prediction of inflation for period $t+1$ at period $t$. This type of models are important for forecasting, especially in our framework where some agents are backward-looking and also rational agents have to incorporate this into their forecasts. Thus we estimate the model (9) that is stated in terms of observable variables with the restrictions on all coefficients, where $\eta_0$ and $\eta_1$ are REE coefficients. Our formulation is inherently different than the one by Carroll (2003a, 2003b) as epidemiological framework that he proposes is no longer valid in our setup where subjects in principle observe all relevant information.[21] About 97% of agents display a significantly positive $\lambda_1$, while the average $\lambda_1$ is 0.20. Groups in treatment 3 had the highest average $\lambda_1$ (0.37), while subjects in treatment 2 had the lowest (0.11). It is not straightforward to define rationality in our framework and thus the results can be challenged on these grounds. The definition used in this subsection corresponds to REE if all agents in the group form expectations rationally.[22] Similar weighted average regressions are estimated in Heemeijer, Hommes, Sonnemans, and Tuinstra (2007), where they replace RE prediction with the equilibrium price.

## 4.3 Trend Extrapolation Rule

We also evaluate simple trend extrapolation rules. These are pointed out as particularly important rules for expectation formation process in Hommes, Sonnemans, Tuinstra, and

[20] In field experiments by Berlemann and Nelson (2005) similar rationality tests were conducted suggesting that most participants exploit all available information.

[21] He argues that news about inflation spreads slowly across agents and reaches only a fraction $\lambda_1$ of population in each period.

[22] Note if we would use naive expectations this model would correspond to adaptive expectations in equation (11).

van de Velden (2005). We specify the following process:

$$\pi^k_{t+1|t} - \pi_{t-1} = \tau_0 + \tau_1 \left(\pi_{t-1} - \pi_{t-2}\right), \tag{10}$$

where we estimate $\tau_0$ and $\tau_1$. We find that constant is significant at 5% level in 28.7% of cases while the $\tau_1$ is significant in 78.2% of cases at the same level. Most of the times $\tau_1$ is between 0 and 1, but there are a few cases when $\tau_1$ is significantly lower than 0 (6.9%) and for 15.3% of subjects it is significantly higher than 1. We refer to the latter rules as strong trend extrapolation. Hommes, Sonnemans, Tuinstra, and van de Velden (2005) find that about 50% of subjects in their experiment behave consistently with the trend extrapolation rule.

## 4.4 Estimating Simple Learning Rules

In order to test for adaptive behavior, we apply different learning rules to experimental data. For a discussion on learning rules and convergence to rational expectations see Evans and Honkapohja (2001). We first test learning on a model with constant gain updating (CGL), where subjects learn from their past observed errors. The model below is equivalent to the adaptive expectations formula:

$$\pi^k_{t+1|t} = \pi^k_{t-1|t-2} + \vartheta \left(\pi_{t-1} - \pi^k_{t-1|t-2}\right), \tag{11}$$

where $\vartheta$ is the constant gain parameter. Under this learning rule agents revise their expectations according to the last observed error. In the experiment subjects are asked to forecast inflation in the next period (hence they make their forecast for period $t+1$ at time $t$), therefore the revision regards their previous period's forecast $(t-1)$, which is made at time $t-2$. Note that this rule corresponds to the second order adaptive scheme in Marimon, Spear, and Sunder (1993). All participants have $\vartheta$ positive and significant at a 5 percent level. 13.4% of participants have a constant gain parameter significantly lower than 1, while 53.7% of them update their forecasts with an error correction term significantly greater than 1. This means that the latter agents possibly overreact to their past errors. Their prevalence might imply problems with dynamic stability in certain treatments.

Below we present a learning mechanism with decreasing gain parameter (DGL):

$$\pi^k_{t+1|t} = \pi^k_{t-1|t-2} + \frac{\iota}{t} \left(\pi_{t-1} - \pi^k_{t-1|t-2}\right). \tag{12}$$

If the estimated parameter ($\iota$ in this version) is significantly different from 0, we conclude that agents actually learn from their past mistakes with a decreasing gain over time. Our tests do not support the hypotheses that the coefficient decreases over time as the $R^2$ is always greater (for all subjects) for a constant gain model.

Several versions of these models are estimated in Arifovic and Sargent (2003), Hommes, Sonnemans, Tuinstra, and van de Velden (2005), Marimon and Sunder (1995) and Bernasconi and Kirchkamp (2000). Hommes, Sonnemans, Tuinstra, and van de Velden (2005) argue that some subjects (about 5%) behave consistently with this rule, while Marimon and Sunder (1995) and Bernasconi and Kirchkamp (2000) put forward that most subjects in their OLG experiments use either first or second order adaptive expectations.

### 4.4.1 Recursive Representation of Simple Learning Rules

The above specification mainly aims at testing whether data support the existence of adaptive behavior. As in the adaptive learning literature in this subsection we assume that subjects behave like econometricians, using all available information at the time of the forecast. In the following specifications, we test whether agents update their coefficients with respect to the last observed error. We assume four different perceived laws of motion (PLM):

$$\pi^k_{t+1|t} = \phi_{0,t-1} + \phi_{1,t-1}\pi_{t-1} + \varepsilon_t. \tag{13}$$

$$\pi^k_{t+1|t} = \phi_{0,t-1} + \phi_{1,t-1}y_{t-1} + \varepsilon_t. \tag{14}$$

$$\pi^k_{t+1|t} = \phi_{0,t-1} + \phi_{1,t-1}\pi^k_{t|t-1} + \varepsilon_t. \tag{15}$$

$$\pi^k_{t+1|t} - \pi_{t-1} = \phi_{0,t-1} + \phi_{1,t-1}\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t. \tag{16}$$

Note that equation (14) represents a PLM of the REE form and equation (16) a version of the trend extrapolation rule. When agents estimate their PLMs they exploit all available information up to period $t-1$. As new data become available they update their estimates according to a stochastic gradient learning (see Evans, Honkapohja, and Williams, 2009) with a constant gain. Let $X_t$ and $\widehat{\phi}_t$ be the following vectors: $X_t = \begin{pmatrix} 1 & \pi_t \end{pmatrix}$ and $\widehat{\phi}_t = \begin{pmatrix} \phi_{0,t} & \phi_{1,t} \end{pmatrix}'$. In this version of constant gain learning (CGL) agents update coefficients according to the following rule:

$$\widehat{\phi}_t = \widehat{\phi}_{t-2} + \vartheta X'_{t-2}\left(\pi_t - X_{t-2}\widehat{\phi}_{t-2}\right). \tag{17}$$

The empirical approach consists in searching the parameter $\vartheta$ that minimizes the sum of squared errors (SSE), i.e. $\left(\pi^s_{t+1|t} - \pi^k_{t+1|t}\right)^2$ (see Pfajfar and Santoro, 2010 for details). The implicit problem in this approach is that we have to assume the initial values for $\widehat{\phi}_t$

for 2 periods. Setting up the initial values is one of the main problems when we recursively estimate learning. This issue is extensively discussed in Carceles-Poveda and Giannitsarou (2007). Strictly speaking, this problem should not occur in our case since we simply try to replicate our time-series data as closely as possible. Thus, in the following recursive learning estimations, we design an exercise in order to search for the best combinations of the gain parameter and initial values to match each subjects' expectations as closely as possible. This strategy can also be considered as a testing procedure for the detection of learning dynamics for each individual. If the gain is positive under this method of initialization, then the series would exhibit learning for all other initialization methods with higher (or equal) gain.

We find that 56.5% of participants learn according to the first setup with lagged inflation as in model (13). The gain parameter $\vartheta$ is in the range between 0.0001 and 0.1000, with a mean value of 0.02900 and the median is 0.01125. We also estimate adaptive learning with the PLMs consistent of the REE form and AR(1) form, however these models rarely outperform other models studied here. In the learning version of the trend extrapolation model (16) 31.5% of subjects have positive gains. The optimal gains are on average slightly higher than before as they range between 0.0003 and 0.7900 with a mean value of 0.0654 (the median is 0.0310).

This version of the PLM (16) often performs better than previous versions of learning in terms of SSE. Below we compare different models and find that this version of constant gain learning indeed best represents the behavior of a significant proportion of our subjects. For a comparison with other studies, we exclude from our sample all subjects for which learning does not represent the best model.[23] In this case, we find that the average gain of these subjects is 0.0447 with a standard deviation of 0.0537 (median is 0.0260). The standard deviation is quite high as there are a few very high values, but most of the gains fall in the range between 0.01 and 0.07.

There are only a few estimates of the gain coefficient in the literature. Orphanides and Williams (2005a) suggest a gain between $0.01-0.04$ and Milani (2007) estimates it at 0.0183, while Pfajfar and Santoro (2010) find smaller gains (around 0.00021 for a similar version of learning). Results in this paper suggest slightly higher gains than most of the above papers, but our data might be more volatile than the actual US inflation. However, we have to point out that estimates in this paper represent the first results on "clean" inflation data obtained from individuals.

[23] We will consider Comparison 1 in the Table 5 and exclude model (14) as it is generally associated with extremely high values of gain parameter.

## 4.5 "General" Models of Expectation Formation

Simple learning rules do not capture all macroeconomic factors that can affect inflation forecasts. In this subsection we estimate some general models of expectation formation. We specify the following regression:[24]

$$\pi^k_{t+1|t} = \alpha + \gamma\pi_{t-1} + \beta y_{t-1} + \mu i_{t-1} + \zeta\pi^k_{t|t-1} + \varepsilon_t. \tag{18}$$

We find that 81.9% of agents take into account inflation when making their predictions. About 56.0% of the subjects take interest rate into account, while 66.7% also regard their own forecast from the previous period. Under some restrictions this equation could represent the form of the RE solution of the model ($\zeta = 0$).[25]

For a comparison we also estimate a simple AR(1) model:

$$\pi^k_{t+1|t} = \phi_0 + \phi_1\pi^k_{t|t-1} + \varepsilon_t. \tag{19}$$

Similar model was already estimated with recursive learning. Model with constant coefficients, in general, is not often used by subjects for forecasting inflation.

## 4.6 Comparison with "Classical Econometrician" and Rational Expectations

Before we discuss the best performing model for each individual we ask ourselves how would a "classical" econometrician forecast inflation in this environment.[26] We estimate a regression for each period in time using only the available information that is on the subjects' screens. Therefore, we estimate rolling regressions and make one-step-ahead forecasts. Similar approach is used by Branch (2004) in the survey data literature for proxying the rational expectations. He uses a trivariate VAR model and estimates it recursively. In our case,

[24] Models in groups 19-24 do not have interest rate as dependent variable as this would imply multicolinearity due to the design of monetary policy in our framework.

[25] We also investigate more in depth the nature of the forecast error. We estimate the model where we regress the forecast error on past observed forecast error and changes of other macroeconomic variables. Subjects often do not exploit the informational content of the output gap and most importantly subjects overreact to last observed change of inflation. As the coefficient in front of the change in inflation is in most cases higher than 1, we can say many subject are pessimistic about future developments of inflation. This feature is repeatedly found in the survey data literature.

[26] Of course, a more "sophisticated econometrician" could do a better job. For example, exogenous shocks are not observable in our framework, but a better econometrician could design an unobserved components model to extract information about the autoregressive shocks and then use them in these regressions. In the RE paradigm shocks play a significant part in the formation of expectations. In some treatments it is possible to observe that at least some agents extract information about the shock in the PC and at least partly use this information when forecasting. This is especially evident in treatment 4.

because of degrees of freedom problems, we have to resort to a univariate model (18). For a comparison, we also recursively estimate the adaptive expectations model (11) and a version of trend extrapolation rule without the restrictions on coefficients. In practice, this rule is then equivalent to the AR(2) model. We evaluate the general model with and without the restriction: $\zeta = 0$. Then we compare these forecasts with the actual realizations and compute SSE, which are presented in the table below for five competing models. Before starting the analysis it is worth pointing out that in treatments where the variance of inflation is higher also the mean SSE is higher (correlation coefficient is 0.91). In two thirds of our groups the trend extrapolation rule is performing best. However, in more stable treatments the general model can also outperform the trend extrapolation rule.

Insert Table 4 about here

This table gives us a benchmark for evaluating prediction accuracies of subjects. It is interesting to note that best performing subjects often outperform our classical econometrician (best performing model). This practically occurs in all groups except the ones comprising treatment 3 where high frequency of cycles is observed (see Figure 2). There are two possible explanations for that: first, some subjects are actually rational or at least forward-looking and second, subjects switch between different expectation formation mechanisms. We start by investigating the first possibility while in the next section we dig deeper regarding the second possible explanation.

There are two definitions of rationality: the statistical and "economic" definition. The former is defined and discussed in section 4.1, while the latter interpretation argues that expectations should be consistent with the underlying economic model. Strictly speaking, where all agents know the macroeconomic model and behave accordingly, we know exactly the form of RE and the actual coefficient values.[27] However, in our experiment subjects are not familiar with the underlying macroeconomic model, and they might reasonably believe that other subjects potentially do not use RE. They have to take this into account when producing inflation forecasts. Even more, if rational agents understand the informational content of the interest rate, especially in treatments 1-3, they could implement this information into their decisions. Thus, in the environment of heterogeneous forecasts the REE PLM may be of a different form than the REE PLM in the case of homogeneous forecasts. This issue is further discussed in Nunes (2009) and Molnár (2007) where it is conjectured that some proportion of agents use adaptive learning to forecast, while the remaining agents are rational. Nunes (2009) studies this problem in the context of forward-looking NK model and shows how to solve the model under the assumption of heterogeneous expectations. Our case is slightly

[27] As it can be seen below, this REE PLM model (14) never outperforms other models.

different as the information sets of individuals do not include other subjects' forecasts. These could only be observed indirectly through interest rate in treatments 1-3, however subjects do not know that the interest rate setting depends on their forecasts. Nevertheless, if some agents use the PLM with last observed inflation, and the rational agents are aware of that, then they have to include the last observed inflation in their PLMs as well. As there is not possible to calculate RE as a benchmark in our heterogeneous environment we have two different possibilities: (i) to use the statistical definition of rational expectations mentioned above, or (ii) to estimate the ALM (actual law of motion) for inflation in each group and check whether the estimated coefficients of the corresponding PLM entail statistically different coefficients to the ones of ALM. The problem here is that it is not straightforward how to define the form of the ALM as discussed above. We assume that the ALM is of the following form:

$$\pi_{t+1} = \gamma_0 + \gamma_1 \pi_{t-1} + \gamma_2 \pi_{t-2} + \gamma_3 y_{t-1} + \gamma_4 i_{t-1} + \varepsilon_t, \tag{20}$$

and the corresponding correctly parameterized PLM is:

$$\pi^k_{t+1|t} = \beta_0 + \beta_1 \pi_{t-1} + \beta_2 \pi_{t-2} + \beta_3 y_{t-1} + \beta_4 i_{t-1} + \varepsilon_t. \tag{21}$$

In order that we can claim that one subject has model consistent or RE the estimated coefficients in both regressions should not be statistically different. To test for that we estimate the following equation:

$$\pi_{t+1} - \pi^k_{t+1|t} = \mu_0 + \mu_1 \pi_{t-1} + \mu_2 \pi_{t-2} + \mu_3 y_{t-1} + \mu_4 i_{t-1} + \varepsilon_t, \tag{22}$$

where $\mu_i = \gamma_i - \beta_i$. For subject to forecast rationally all estimated coefficients (jointly) in equation (22) should not be statistically significant. In the discussion below we compare these definitions of RE. Rationality is in this case "superimposed" as we classify all agents that satisfy the requirements as rational irrespective of their expectation formation mechanism.

## 4.7 Discussion

In this section we determine which theoretical model on average best describes the behavior of each individual. We compare the SSE[28] of each individual for the 10 models of expectation formation that are described above. A subject is regarded to use the model which produces the lowest SSE between the model predictions and their actual predictions.

We compare 9 models of inflation expectation formation that best describe the behavior

[28] Results and conclusions are the same irrespectively whether we use RMSE (root mean square error), $R^2$ or SSE as they are all monotonic transformations of each other.

of at least 1 participant. Model (12) is never used as it is always outperformed by other models.

Insert Table 5 about here

In Table 5 we present 6 different comparisons as we use different definitions of the RE. In comparisons 1 and 2 we define the RE based on statistical properties while in comparisons 3 and 4 based on theory as it is outlined above in section 4.6 in equation (22): in comparison 1 (3) at 5% significance level and in comparison 2 (4) at 1% significance level. In comparison 6 we compare all empirical models, while in comparison 5 we exclude the general model from the set of alternative models.

We can observe that results are indeed quite similar across alternative definitions of RE, although the theoretical definition (comparisons 3 and 4) suggests a slightly higher proportion of rational subjects. One possible reason is that we estimate the model (22) under the assumption of common AR(1) errors as the experiment design embeds unobserved AR(1) shocks. Without this assumption comparisons 3 and 4 would imply 27.3% and 31.0% of rational subjects. Generally, there is evidence that in all treatments about $30-45\%$ of subjects are rational and about $25-35\%$ of agents simply extrapolate trend. Around $5-10\%$ of agents employ adaptive expectations while the remaining $15-25\%$ of subjects mostly behave in accordance with new theories of expectation formation, adaptive learning and sticky information type models.

As mentioned before, most of other papers in the experimental literature stress the importance of adaptive expectations. Expectation formation of prices is also studied in the US beef market where Baak (1999) and Chavas (2000) show evidence of heterogeneity. Chavas (2000) estimates that 81.7% of agents are boundedly rational using simple univariate models to forecasts prices. The remaining 18.3% of agents are rational. Baak (1999) uses different techniques and finds that the proportion of rational agents is higher. He estimates that about two thirds of agents are rational, while others are boundedly rational.

The remaining literature focuses on the analysis of survey data. Branch (2004) presents the results for 3 competing models of expectation formation (VAR, adaptive, and naive) estimated based on Michigan survey data. He finds that about 48% of agents use a VAR predictor and 44% of agents behave adaptively, while the naive predictor accounts for the remaining 7% of the sample. Especially our comparisons 2 and 4 yield very similar estimates. In the same vein, although using different techniques, Pfajfar and Santoro (2010) analyze how agents form expectations in the Michigan Survey. They also test for adaptive learning and sticky information models. It is found that about 44% of agents use these new models of expectation formation, while there are only about 7% of agents that behave rationally.

Around 10% of agents are found to be static and the remaining 39% use simple univariate rules to forecast, where they predominantly rely on either their past forecasts or the past value of inflation.

The availability of information is probably the main reason why our results on rationality are different from some previous studies on the inflation expectation process in survey data. We must bear in mind that subjects in our experiment have always available historical series on all relevant macroeconomic variables and their past predictions. In the real world all variables might not be readily observable or the information cost for collecting them might play an important role. The other reason for high degree of rationality is that we initialize the model under RE and that we have "pure" data on inflation expectations. This increases the possibility of not rejecting the assumption of rationality.

We further study the degree of heterogeneity by analyzing each treatment separately. We present comparison 1 across all treatments in Table A1 in Appendix A where we can observe that there is quite a lot of heterogeneity across treatments. We further discuss this in the next section, where we analyze switching between different rules.

# 5 Switching Between Different Models

The aim of this section is to further investigate how subjects form expectations. Do they consistently use one model or do they switch between different models? We mentioned before that switching might be one of the explanations for better performance of some individuals compared to the "classical econometrician." There are some attempts in the literature to link the performance of forecasting rules to the share of agents using that rule. Models that explore this issue are generally labelled as rationally heterogeneous expectations models. Some examples of these models are Brock and Hommes (1997), Branch and McGough (2007) and Pfajfar (2008). Their main argument is that it is not always optimal from an utility maximization point of view to forecast rationally as this might entail some costs.

In this section we tackle the problem from a slightly different perspective as we only have 9 subjects in each group. Their information sets are different as subjects do not directly observe past forecasts of other subjects. Thus it is not possible to compare these different models of dynamic predictor selection in our setup. We rather focus on establishing some stylized facts about "unrestricted" switching on the individual basis. Alternative approach, where all agents have the same information set is investigated in Anufriev and Hommes (2008). They provide support for switching based on a version of the predictor dynamics analyzed in Hommes, Huang, and Wang (2005) and show that in an asset pricing environment the model with switching between simple heuristic rules can replicate the main results of the

Hommes, Sonnemans, Tuinstra, and van de Velden (2005) experiment in terms of individual behavior and aggregate dynamics. We proceed this analysis somehow differently as our results above postulate that many of the employed rules are based on personal information, i.e. they include their own past forecast (which is unobservable to others) to their forecasting rule. In essence, we look at the roots of the switching behavior, where we do not impose a particular switching mechanism.

## 5.1 Unrestricted Switching

We start this analysis by determining the optimal model for each individual in each period with a recursive estimation of the models specified above. Our approach consists of recursively computing the SSE up to a period $t$ and then compare them in a period $t$ for each individual. This comparison is performed for all periods except for the first 4 periods. Therefore, we can determine which model best fits the actual forecasting series in each point in time and whether there is any switching observed among these models. As many models' predictions are very similar at least in some episodes, we assume that there is no switching if the model that performs best in the previous period is not outperformed in the current period by 0.1 percentage points in terms of forecasts accuracy or 0.01 in terms of SSE. The rationale behind this choice is that the majority of forecasts are reported to one decimal point accuracy and subjects are not able to differentiate between these competing models. The relative shares of each model are reported in Table 6.

Insert Table 6 about here

We can observe that higher proportion of all forecasts are made using one of the stochastic gradient learning algorithms. Depending on the treatment, in 23 to 45% of all cases agents use these algorithms to forecast. If we average this across groups, 36.7% of the forecast decisions are best explained with adaptive learning. This means that, on average, adaptive learning is the most popular way of forming beliefs.

In around 17% of cases subjects use the general model, and in about 12% of all forecast decisions they behave in accordance with the sticky information type model. The remaining third of all forecasts are best explained with some sort of backward-looking models. In specific, around 14% of cases subjects use simple trend extrapolation rules while the remaining 20% of cases they behave in an adaptive manner. Compared to the results outlined above for "average" best model, we can immediately observe that there is approximately the same proportion of backward-looking cases as there are subjects that use backward-looking rules. However, when allowing for switching there are more forecast decisions made in an adaptive way. Also model (15) is only a predominant model for one subject, but when we allow for switching it is used on average in 15.6% of all forecasts.

Generally, we can observe that when we allow subjects to switch between different models, they are in fact using alternative models to forecast. Under this assumption, agents use between 1 and 7 different models (average number of models used for forecasting is 6.5) and they on average switch every 4 periods. However, switching is occurring less frequently in treatments 3 and 4 compared to treatments 1 and 2 (significant at 5% level with different tests of equality of medians).[29] Only one subject did not switch between models. Overall, these results support the idea of intrinsic heterogeneity that is theoretically modelled in Branch and Evans (2006) and Pfajfar (2008).

To further analyze the degree of heterogeneity in the data, we compute the average number of models used in each period. We find that on average 4.5 different models (between 2 and 7) are used within a group in each period. This additionally supports the above conjecture that heterogeneity is pervasive as there are not significant differences across treatments. The average number of models employed for forecasting within a group varies (in each period) only between 4.2 and 5.3. Furthermore, there is no "smoothing" employed across different subjects in the same group. We have only employed some "smoothing" within each subject as some models perform quite similarly and cannot be differentiated at one decimal point accuracy.

We also investigate the pattern (timing) of switching with panel probit and logit models (with random, and fixed effects, and population averages), where dependant variable, $z_t^k$, is 1 when switching occurs and 0 otherwise. We estimate the following regression:

$$z_t^k = \alpha_1^k + \alpha_2 \pi_{t-1} + \alpha_3 y_{t-1} + \alpha_4 i_{t-1} + \alpha_5 \left(\pi_{t-1} - \pi_{t-1|t-2}^k\right)^2 + \varepsilon_t^k. \tag{23}$$

We find that subjects decide to switch according to developments of inflation, output gap, and interest rate. Alternative models exhibit similar effects of the explanatory variables. The most pronounced effect expectably comes from the output gap which has a strong negative impact on the probability of switching. Positive change in inflation trend increases the probability of switching, however, higher inflation decreases it. This demonstrates that there exists a certain pattern in the structure of individual switching. There are also some differences across treatments, especially in treatment 4 the pattern of switching is different. However, treatment dummies are insignificant if we insert them to the above regression.

[29] With e.g. Kruskal-Wallis test. Switching is occuring on average every 6.1 periods in treatment 4, 3.7 period in treatment 3, 2.6 period in treatment 2, and 2.9 periods in treatment 1.

Results are reported in Table 7.[30]

Insert Table 7 about here

# 6 Monetary Policy in the Presence of Heterogeneous Expectations

Woodford (2003) showed that in this environment monetary policy should minimize variance of inflation and output gap as this corresponds to maximizing utility of consumers. Therefore we start this section with the analysis of variance of inflation as the monetary authority cares only about inflation in instrument rules under scrunity. Tests for differences in medians across treatments where the null hypothesis that the medians are the same in all treatments is rejected at 1% significance with Kruskal-Wallis and van der Waerden tests (see Conover, 1999). Therefore, we can argue that the design of monetary policy matters in our framework. The following table shows the comparison of median standard deviations of inflation in treatments 2, 3, 4 with treatment 1. We report p-values of the Kruskal-Wallis test.[31]

Insert Table 8 about here

We also find that there is a significant difference between treatments 2 and 3 (p-value is 0.0250). Thus, we can argue that treatments 3 and 4 produce significantly lower inflation variability than treatments 1 and 2. Now that we establish that there is a difference in variance of inflation between treatments we further analyze the roots of these differences between and within treatments.

To have an illustration how important are expectations for the stability of the system we simulate our treatments with different forecasting rules under the assumption of homogeneous expectations (see Figures A4 and A5). We can immediately observe that adaptive expectations (with a gain coefficient higher than 1) and trend extrapolation rules can lead to pronounced cyclical variability of inflation. It is also possible to observe that treatments 2 and 4 perform better than 1 and 3 in stabilizing those expectation formation mechanisms. However, the evidence might be reversed with respect to "stable" expectation formation mechanisms.

The proportion of backward-looking (especially trend extrapolation) agents plays a particularly important role for the stability of the system. We can observe that there is a

[30] In this case it is not straightforward whether in the estimation procedure for the standard errors to allow for intragroup correlation or intratreatment correlation. In the main text we report standard errors that are clustered in groups and in the Table A2 in the Appendix A standard erros that are clustered in treatments.

[31] Other nonparametric tests perform very similarly.

considerable degree of heterogeneity across treatments. Even more, differences in the degree of backward-looking subjects can explain the differences in variability between groups in the same treatment. The results are intuitive as we find that there is a strong correlation between the stability of the system and the degree of trend extrapolation behavior. We further test these conjectures regarding the relationship between the variability and proportion of different groups of subjects with cross-sectional and panel data regressions. With former we find that especially increasing proportion of trend extrapolation behavior is increasing the variance. Also increasing proportion of CGL adaptive expectations rules is increasing the variance as most of the estimated coefficients $\vartheta$ in equation (11) are higher than 1 while the proportion of recursive learning (15) and also sticky information rules (8) is reducing it. We estimate the following regressions:

$$sd_s = \eta_0 + \eta_1 p_{js} + \varepsilon_s,$$

where $sd_s$ is standard deviation of group $s$, and $p_{js}$ is proportion of agents using $j$-th model for forecasting in group $s$. The set of alternative models is the same as in Table 6 above. Regression results are reported in Table 9, both with robust and clustered standard errors. Initially, we added treatment dummies to the above regression, however they were insignificant in almost all cases. We have to point out that all estimated coefficients (that are significantly different than 0) have the expected signs.

Insert Table 9 about here

These results are confirmed also with the system GMM estimator of Blundell and Bond (1998) for dynamic panels. To construct the panel we compute the $sd_{s,t}$, standard deviation from the first period up to period $t$. Using the switching analysis we similarly compute $p_{js,t}$, the share of model $j$ in group $s$ up to the period $t$. We estimate the following model:

$$sd_{s,t} = \eta_0 + \eta_L sd_{s,t-1} + \sum_j \eta_j p_{js,t} + \varepsilon_{st}.$$

Results are reported in Table 10. The only intriguing result is about the coefficient on the proportion of the general model (18) which is insignificant in the cross sectional regression and significantly positive or insignificant in dynamic panel data models. Therefore it is difficult to say from this analysis what is the effect of the proportion of usage of general model (18) to the stability of inflation. Although these agents use all relevant information to forecast inflation simulation exercise shows that at low values of $\gamma$ this forecasting model (if used exclusively) will result in high variability of inflation (see Figure A6). Furthermore, theoretical analysis shows that as soon as one uses past inflation to forecast the model exhibits indeterminacy, i.e. there might be a multiple equilibria problem.[32]

Insert Table 10 about here

The result regarding the influence of the proportion of trend extrapolation rules to the standard deviation of inflation is very robust across these different techniques as the coefficients are always very significant and positive. The proportion of these agents probably plays the most important role for the stability of inflation. It also helps us to explain the differences among groups within the same treatment. Generally, we note that the group with lower proportion of trend extrapolation rules is more stable compared to other groups in the same treatment.

Heemeijer, Hommes, Sonnemans, and Tuinstra (2007) compare experimental results in positive and negative expectations feedback models.[33] In a positive expectations system, e.g. asset pricing model, they observe similar aggregate behavior to ours and note that when there is a stronger positive feedback more agents resort to backward-looking, especially trend following rules. In our case, by changing the monetary policy, we augment the degree of positive feedback from inflation expectations to current inflation. Therefore, the design of monetary policy is important for the prevailing expectation formation mechanism and vice versa, as can be seen if we compare results within the same treatment. The graphical analysis of the evolution of inflation across treatments is reported in Figure 2.[34]

Insert Figure 2 about here

However, this is only a part of the story in our experiment. We expected that the treatment 2 where monetary authority does not react too strongly to inflation expectations ($\gamma = 1.35$) performs better regarding the stability of inflation than the benchmark treatment, although the theory under rational expectations suggests that higher $\gamma$ leads to lower variability of inflation. This is not confirmed in our analysis above as the median standard deviation is not statistically different than in treatment 1. This might be due to expectations of cycles by some individuals in groups 4 and 5 of this treatment and extensive use of strong trend extrapolation rules at the beginning of the experiment. In order to study the relationship between $\gamma$ and the variance of inflation under different expectation formation mechanisms we design simulation exercises that exactly replicate the design, parametrization and shocks employed in the experiment. When all subjects have rational expectations

[32] Proof of this statement can be obtained from the authors upon request.

[33] Also Fehr and Tyran (2008) compare the two environments, although in a different context.

[34] Detailed Figures with the evolution of inflation and inflation forecasts in each treatment are reported in Appendix A (Figures A2 and A3).

we confirm the theory that higher $\gamma$ leads to lower variability of inflation while many other expectation formation mechanisms produce non-monotonic, often U-shaped behavior of the inflation variance. On one hand, rules that we labelled as stable in regressions above produce decreasing variability of inflation when increasing $\gamma$, although sometimes non-monotonic. On the other hand, especially trend extrapolation rules will lead to U shaped behavior and eventually higher variability when increasing $\gamma$ (see Figure A6). The minimum variability of inflation with sticky information and trend extrapolation rule is achieved at $\gamma = 1.1$. For naive expectations the minimum is around $\gamma = 3$ (non-monotonic U-shaped). This can be also observed from Figures A4 and A5. Therefore, the relationship between the variability of inflation and different rules is nontrivial and the question whether treatment 2 should produce lower variability compared to treatment 1 depends particularly on the proportions of alternative rules used. Based on simulation results and observed behavior of individuals we can argue that in the presence of heterogeneous expectations instrumental rules that are less aggressive have the potential to produce lower variability of inflation, however there is the risk that, e.g. after a shock, the amplitude of inflation increases significantly as monetary policy is not aggressive enough. Thus, one could argue that non-linear Taylor-type rules would perform best in this environment, although the literature in monetary economics has not attached much attention to this type of instrumental rules.

As we have seen so far, the expectational feedback is not the only source of instability in our multivariate system where we also have lagged endogenous variables.[35] Treatment 3 produces lower variability of inflation compared to treatment 2, but in the former case the frequency of cycles is significantly higher as the monetary authority is (too) strongly responding to deviations from inflation target. After some threshold of response to inflation forecast (depends on the proportion of agents using each rule) the resulting amplitude of the inflation variability decreases, while the frequency of cycles increases. The latter makes it more difficult to forecast and more participants resort to simpler rules. Using simulations explained above we can identify two effects of increasing $\gamma$ on the variability of inflation: (i) this always increases the frequency of cycles irrespective of the expectation formation mechanism and (ii) it is increases or decreases the amplitude of the cycle. The latter result depends on the expectation formation mechanism and can produce non-monotonic or even U-shaped responses of variability, except for rational expectations where it decreases monotonically (see Figure A6).

Also treatment 4 performs better than the benchmark treatment. Responding to contemporaneous inflation (as in treatment 4) turns out to be a better practise for central banks

[35] Generally, as $\gamma$ is increasing the positive feedback is decreasing.

compared to responding to inflation expectations.[36] Moreover, this treatment resembles quite closely the behavior of survey forecasts, as there are periods when subjects systematically overpredict inflation (low and stable inflation) and underpredict inflation (when inflation is high). This is evident in Figure A3 in the appendix A. In this treatment there is the highest proportion of biased agents and also results from the general model suggest similar behavior of these agents to the results obtained in the survey data literature. Moreover, if we compare the means of inflation forecasts in treatments 1 and 4 we find that the mean of inflation forecasts of groups in treatment 4 is significantly higher than the mean of inflation forecasts of groups in treatment 1 (at 10% significance with Kruskal-Wallis test). Also average inflation in treatment 4 is higher (3.10 in treatment 4 compared to 3.00 in treatment 1), however the difference is statistically insignificant with nonparametric tests. Comparison between treatments 1 and 4 implies that significantly lower standard deviation of inflation (and inflation forecasts) for treatment 4 (see Table 8) comes at a "cost" of higher inflation expectations (and possibly inflation). This result is similar to Bernasconi and Kirchkamp (2000) as they suggest Friedman's money growth rule produces less inflation volatility, but higher average inflation compared to constant real deficit rule.

We can also observe that generally the variability of inflation is lower than the variability of inflation expectations. This provides an explanation to the fact that responding to current inflation stabilizes the system in a more efficient way compared to reacting to expected inflation. Moreover, by reacting to current inflation we decrease the expectational feedback compared to responding to the expected inflation. As a result, in treatment 4 we reduce the size of the expectational cycles as in booms monetary policy overreacts less than in the case when interest rate is set to respond to expected inflation (in presence of backward-looking agents). At the root of this pattern is that backward-looking subjects do not observe the informational content of output gap and do not predict the change in the growth rate of inflation. They still expect that inflation will accelerate as in the last few periods. Then, if the monetary authority is reacting with respect to the expected inflation, they do not change the stance of monetary policy in time. The economy is pushed in the recession where the backward-looking agents underpredict inflation and the recession is more severe than if all agents were rational. The whole process repeats in the next cycle. We have to point out that the causality goes in both directions as the proportion of backward-looking agents (especially strong trend extrapolation agents) depends on the design of monetary policy (degree of aggressiveness) and also the stability of the economy is influenced by the degree

[36] Pfajfar and Santoro (2008b) and Muto (2008) reach similar conclusion in different versions of the NK model: Muto (2008) in case when agents learn from central banks' forecasts, while Pfajfar and Santoro (2008b) when they introduce the cost channel and capital market imperfections.

of backward-looking agents.

Adam (2007) obtains similar dynamic pattern of inflation and inflation expectations, especially to our treatment 3. He argues that the cause for observed behavior is the subjects' reliance to simpler underparametrized rules for forecasting inflation. Thus, he characterizes the dynamics of inflation as a restricted perception equilibrium, as inflation exhibits excessive volatility around its REE. Our paper supports his findings as some agents do not take into account output gap when forecasting. However, we also show that the volatility of inflation depends on the way monetary policy is designed and conducted. We argue that the proportion of backward-looking subjects plays an important role, especially those that use strong trend extrapolation rule.[37]

# 7 Conclusion

In this paper we design a macroeconomic experiment where subjects are asked to forecast inflation. The underlying model of the economy is a simple NK model which is commonly used for the analysis of monetary policy. The focus in this paper is on the formation of inflation expectations and monetary policy design. In different treatments we employ various modifications of the original Taylor rule and study the influence of alternative monetary policy designs to inflation formation process and also vice versa. Therefore, we also try to determine the design of monetary policy which would effectively stabilize and anchor the process of inflation expectations. It is clear that monetary policy influences the expectation formation process. We find that the variability of inflation is significantly lower in treatments 3 and 4 compared to treatments 1 and 2. The cyclical behavior of inflation is also studied in the experimental study by Adam (2007). When we set interest rate with respect to current inflation, we observe the dynamics of inflation expectations that most closely resembles the behavior of survey data. Generally, this setup performs better in terms of inflation variability than responding to the expected inflation as the variability of inflation is lower than the variability of inflation forecasts. Thus, we reduce the amplitude of expectational cycles.

However, we can point out that the underlying process of inflation expectation formation depends also on the way monetary policy is conducted. The proportion of backward-looking agents, especially trend extrapolating subjects, plays an important role, as in some environments it is more difficult to forecast inflation rationally. In these cases more subjects

[37] Also several asset pricing experiments have observed the dynamics of aggregate price exhibiting bubbles (see eg. Smith, Suchanek, and Williams, 1988 and Hommes, Sonnemans, Tuinstra, and van de Velden, 2005). Even more, Lei, Noussair, and Plott (2001) show that this can occur also in an environment where speculation is not possible. They conclude that this occurs due to systematic errors in decisions.

resort to simpler backward-looking rules. We find that roughly $30 - 35\%$ of subjects predominately use trend extrapolation rules and additionally $5 - 10\%$ of subjects use adaptive expectations. Contrary to previous studies, our results suggest that there is a significant and relatively large share of agents that predominately use rational expectations. The share of these agents is about $35 - 45\%$. The remaining agents use some version of adaptive learning or sticky information type models. Furthermore, we also find that most agents tend to switch between different rules. When we take into account this possibility, we get slightly different results. Most notably, adaptive learning models become more important as this mechanism for forecasting is used in 36.7% of all forecasting decisions. This paper is one of the first empirical contributions to postulate that these models represent one of the most popular ways of forecasting inflation. The average proportion of trend extrapolative decisions is smaller when we allow for switching (14%), but in accordance to our conjecture above it varies significantly across treatments (between 6.1 and 23.6%). In 16.9% of cases agents use the general model, 20.2% adaptive expectations, and the remaining 12.1% of cases agents use sticky information type model.

# References


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

ANUFRIEV, M., AND C. HOMMES (2008): "Evolutionary Selection of Individual Expectations and Aggregate Outcomes," Mimeo, University of Amsterdam.

ARIFOVIC, J., AND T. J. SARGENT (2003): "Laboratory Experiments with an Expectational Phillips Curve," in *Evolution and Procedures in Central Banking*, ed. by D. E. Altig, and B. D. Smith. Cambridge University Press, Cambridge.

BAAK, S. J. (1999): "Tests for bounded rationality with a linear dynamic model distorted by heterogeneous expectations," *Journal of Economic Dynamics and Control*, 23(9-10), 1517–1543.

BAKHSHI, H., AND A. YATES (1998): "Are U.K. Inflation Expectations Rational?," Discussion Paper 81, Bank of England Working Paper.

BERLEMANN, M., AND F. NELSON (2005): "Forecasting Inflation via Experimental Stock Markets Some Results from Pilot Markets," Ifo Working Paper Series Ifo Working Paper No. 10, Ifo Institute for Economic Research at the University of Munich.

BERNANKE, B. (2007): "Inflation Expectations and Inflation Forecasting," Discussion paper, Speech at National Bureau of Economic Research Summer Institute, Cambridge, Massachusetts.

BERNASCONI, M., AND O. KIRCHKAMP (2000): "Why do monetary policies matter? An experimental study of saving and inflation in an overlapping generations model," *Journal of Monetary Economics*, 46(2), 315–343.

BLUNDELL, R., AND S. BOND (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87(1), 115–143.

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

——— (2007): "Sticky information and model uncertainty in survey data on inflation expectations," *Journal of Economic Dynamics and Control*, 31(1), 245–276.

BRANCH, W. A., AND G. W. EVANS (2006): "Intrinsic heterogeneity in expectation formation," *Journal of Economic Theory*, 127(1), 264–295.

Branch, W. A., and B. McGough (2007): "Replicator Dynamics in a Cobweb Model with Rationally Heterogeneous Expectations," *Journal of Economic Behavior and Organization*, forthcoming.

——— (2009): "A New Keynesian model with heterogeneous expectations," *Journal of Economic Dynamics and Control*, 33(5), 1036–1051.

Brock, W. A., and C. H. Hommes (1997): "A Rational Route to Randomness," *Econometrica*, 65(5), 1059–1096.

Bryan, M. F., and S. Palmqvist (2005): "Testing Near-Rationality using Detailed Survey Data," Working Paper Series 183, Sveriges Riksbank.

Carceles-Poveda, E., and C. Giannitsarou (2007): "Adaptive learning in practice," *Journal of Economic Dynamics and Control*, 31(8), 2659–2697.

Carroll, C. D. (2003a): "Macroeconomic Expectations Of Households and Professional Forecasters," *The Quarterly Journal of Economics*, 118(1), 269–298.

——— (2003b): "The Epidemiology of Macroeconomic Expectations," NBER Working Papers 8695, National Bureau of Economic Research, Inc.

Chavas, J.-P. (2000): "On information and market dynamics: The case of the U.S. beef market," *Journal of Economic Dynamics and Control*, 24(5-7), 833–853.

Conover, W. J. (1999): *Practical Nonparameteric Statistics*. Wiley, 3rd edn.

Cooley, T. F., and E. C. Prescott (1995): "Economic Growth and Business Cycles," in *Frontiers of Business Cycle Research*, ed. by T. F. Cooley. Princeton University Press, Princeton.

Curtin, R. (2005): "Inflation Expectations: Theoretical Models and Empirical Tests," Mimeo, University of Michigan.

Duffy, J. (2006): "Experimental Macroeconomics," Discussion paper, in press, to appear in the New Palgrave Dictionary of Economics, 2nd Ed.

Engle-Warnick, J., and N. Turdaliev (2006): "An Experimental Test of Taylor-Type Rules with Inexperienced Central Bankers," CIRANO Working Papers 2006s-05, CIRANO.

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

Evans, G. W., S. Honkapohja, and R. Marimon (2001): "Convergence in monetary inflation models with heterogeneous learning rules," *Macroeconomic Dynamics*, 5(1), 1–31.

Evans, G. W., S. Honkapohja, and N. Williams (2009): "Generalized Stochastic Gradient Learning," *International Economic Review*, forthcoming.

Fehr, E., and J.-R. Tyran (2008): "Limited Rationality and Strategic Interaction: The Impact of the Strategic Environment on Nominal Inertia," *Econometrica*, 76(2), 353–394.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

Giordani, P., and P. Söderlind (2003): "Inflation forecast uncertainty," *European Economic Review*, 47(6), 1037–1059.

Haruvy, E., Y. Lahav, and C. N. Noussair (2007): "Traders' Expectations in Asset Markets: Experimental Evidence," *American Economic Review*, 97(5), 1901–1920.

Hazelett, D., and A. Kernen (2002): "Hyperinflation and seigniorage in an experimental overlapping generations economy," Mimeo, Whitman College.

Heemeijer, P., C. Hommes, J. Sonnemans, and J. Tuinstra (2007): "Price Stability and Volatility in Markets with Positive and Negative Expectations Feedback: An Experimental Investigation," CeNDEF Working Papers 07-01, Universiteit van Amsterdam.

Hommes, C. (2007): "Complexity, Evolution and Learning: a simple story of heterogeneous expectations and some empirical and experimental validation," CeNDEF Working Paper 07, University of Amsterdam.

Hommes, C., H. Huang, and D. Wang (2005): "A robust rational route to randomness in a simple asset pricing model," *Journal of Economic Dynamics and Control*, 29(6), 1043–1072.

Hommes, C., 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.

Ireland, P. N. (2004): "Technology Shocks in the New Keynesian Model," *The Review of Economics and Statistics*, 86(4), 923–936.

Lei, V., C. N. Noussair, and C. R. Plott (2001): "Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality," *Econometrica*, 69(4), 831–59.

Mankiw, N. G., and R. Reis (2002): "Sticky Information Versus Sticky Prices: A Proposal To Replace The New Keynesian Phillips Curve," *The Quarterly Journal of Economics*, 117(4), 1295–1328.

Mankiw, N. G., R. Reis, and J. Wolfers (2004): "Disagreement about Inflation Expectations," *NBER Macroeconomics Annual 2003*, 18, 209–248.

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 (1995): "Deoes a Constant Money Growth Rule Help Stabilize Inflation: Experimental Evidence," *Carnegie - Rochester Conference Series on Public Policy*, 45, 111–156.

McCallum, B. T., and E. Nelson (2004): "Timeless Perspective vs. Discretionary Monetary Policy in Forward-Looking Models," *Federal Reserve Bank of St. Louis Review*, 86(2), 43–56.

Milani, F. (2007): "Expectations, learning and macroeconomic persistence," *Journal of Monetary Economics*, 54(7), 2065–2082.

Molnár, K. (2007): "Learning with Expert Advice," *Journal of the European Economic Association*, 5(2-3), 420–432.

Muto, I. (2008): "Monetary Policy and Learning from the Central Bank's Forecast," IMES Discussion Paper Series 08-E-01, Institute for Monetary and Economic Studies, Bank of Japan.

Nunes, R. (2009): "Learning The Inflation Target," *Macroeconomic Dynamics*, 13(02), 167–188.

Orphanides, A., and J. C. Williams (2005a): "The decline of activist stabilization policy: Natural rate misperceptions, learning, and expectations," *Journal of Economic Dynamics and Control*, 29(11), 1927–1950.

Pesaran, M. H. (1987): *The Limits to Rational Expectations*. Basil Blackwell, Oxford, reprinted with corrections 1989 edn.

Pfajfar, D. (2008): "Formation of Rationally Heterogeneous Expectations," Mimeo, University of Tilburg.

Pfajfar, D., and E. Santoro (2008b): "Credit Market Distortions, Asset Prices and Monetary Policy," Cambridge Working Papers in Economics 0825, Faculty of Economics, University of Cambridge.

——— (2010): "Heterogeneity, learning and information stickiness in inflation expectations," *Journal of Economic Behavior & Organization*, 75(3), 426–444.

Pfajfar, D., and B. Žakelj (2010): "Uncertainty and Disagreement in Forecasting Inflation: Evidence from the Laboratory," Mimeo, University of Tilburg and Universitat Pompeu Fabra.

Roos, M. W., and W. J. Luhan (2008): "Are Expectations Formed by the Anchoring-and-adjustment Heuristic? An Experimental Investigation," Ruhr Economic Papers 0054, Rheinisch-Westfälisches Institut für Wirtschaftsforschung.

Schmalensee, R. (1976): "An Experimental Study of Expectation Formation," *Econometrica*, 44(1), 17–41.

Smith, V. L., G. L. Suchanek, and A. W. Williams (1988): "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, 56(5), 1119–51.

Thaler, R. H. (2000): "From Homo Economicus to Homo Sapiens," *Journal of Economic Perspectives*, 14(1), 133–141.

Walsh, C. E. (2003): *Monetary Theory and Policy*. The MIT Press, second edn.

Williams, A. W. (1987): "The Formation of Price Forecasts in Experimental Markets," *Journal of Money, Credit and Banking*, 19(1), 1–18.

Woodford, M. (1996): "Loan Commitments and Optimal Monetary Policy," *Journal of Monetary Economics*, 37(3), 573–605.

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

# Tables and Figures

| | | |
|---|---|---|
| $\beta = 0.99$ | $\overline{\pi} = 3$ | $\nu = 0.6$ |
| $\varphi = 0.164$ | $\lambda = 0.3$ | $\kappa = 0.6$ |

Table 1: McCallum-Nelson Calibration

| Treatment | Groups | Taylor rule (equation) | Parameters |
|---|---|---|---|
| Inflation forecast targeting (1) | 1-6 | Forward looking (3) | $\gamma = 1.5$ |
| Inflation forecast targeting (2) | 7-12 | Forward looking (3) | $\gamma = 1.35$ |
| Inflation forecast targeting (3) | 13-18 | Forward looking (3) | $\gamma = 4$ |
| Inflation targeting (4) | 19-24 | Contemporaneous (4) | $\gamma = 1.5$ |

Table 2: Treatments

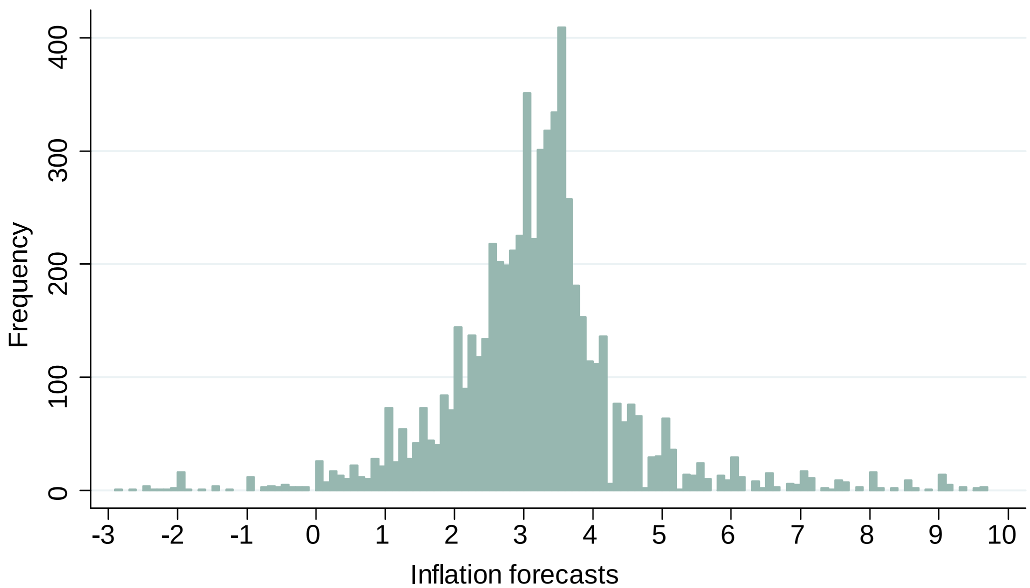


Figure 1: Histogram of inflation forecasts for all treatments.

| Group | 1-1 | 1-2 | 1-3 | 1-4 | 1-5 | 1-6 | 2-1 | 2-2 | 2-3 | 2-4 | 2-5 | 2-6 | 3-1 | 3-2 | 3-3 | 3-4 | 3-5 | 3-6 | 4-1 | 4-2 | 4-3 | 4-4 | 4-5 | 4-6 | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | Inflation Expectations | | | | | | | | | | | | |
| Mean | 2.94 | 3.00 | 3.04 | 3.01 | 3.12 | 3.14 | 3.11 | 3.09 | 3.12 | 3.18 | 2.72 | 3.04 | 3.02 | 3.03 | 3.01 | 3.00 | 3.00 | 3.00 | 3.12 | 3.29 | 3.07 | 3.05 | 3.10 | 3.15 | 3.06 |
| StdDev | 6.31 | 3.48 | 2.02 | 0.73 | 1.12 | 0.93 | 0.75 | 1.87 | 0.49 | 5.77 | 3.76 | 0.86 | 0.57 | 1.07 | 0.26 | 0.30 | 0.30 | 0.23 | 0.38 | 0.89 | 0.48 | 0.36 | 0.54 | 1.42 | 2.20 |
| Min | -13.9 | -6.10 | -2.50 | 0.40 | 0.30 | 0.50 | 1.00 | -0.70 | 0.20 | -12.0 | -8.80 | 0.52 | 1.70 | 0.00 | 2.00 | 1.20 | 2.10 | 2.40 | 2.30 | 1.00 | 1.60 | 2.30 | 0.50 | 0.00 | -13.9 |
| Max | 24.0 | 52.0 | 7.50 | 3.98 | 5.40 | 5.20 | 4.50 | 9.50 | 4.20 | 16.1 | 10.5 | 4.50 | 4.76 | 6.90 | 3.80 | 4.50 | 4.00 | 3.70 | 4.20 | 5.20 | 4.00 | 3.90 | 4.40 | 7.00 | 52.0 |
| | | | | | | | | | | | | | Inflation | | | | | | | | | | | | |
| Mean | 2.85 | 2.88 | 2.92 | 3.00 | 3.13 | 3.12 | 3.12 | 3.09 | 3.13 | 3.02 | 2.52 | 3.03 | 3.01 | 3.02 | 2.99 | 3.00 | 2.99 | 3.01 | 3.09 | 3.23 | 3.05 | 3.05 | 3.09 | 3.11 | 3.02 |
| StdDev | 5.83 | 2.89 | 1.95 | 0.75 | 1.09 | 0.90 | 0.76 | 1.81 | 0.51 | 5.50 | 3.56 | 0.88 | 0.51 | 0.94 | 0.24 | 0.26 | 0.31 | 0.24 | 0.39 | 0.81 | 0.48 | 0.38 | 0.51 | 1.28 | 1.37 |
| Min | -9.53 | -5.27 | -0.84 | 0.67 | 0.81 | 1.16 | 1.26 | 0.06 | 1.84 | -9.04 | -6.74 | 0.80 | 2.00 | 0.97 | 2.49 | 2.41 | 2.48 | 2.51 | 2.40 | 1.77 | 1.88 | 2.46 | 1.77 | 0.68 | -9.53 |
| Max | 16.7 | 10.5 | 6.51 | 3.89 | 5.01 | 4.78 | 4.38 | 7.42 | 3.98 | 12.6 | 8.17 | 4.13 | 3.84 | 5.28 | 3.44 | 3.46 | 3.74 | 3.56 | 3.78 | 4.49 | 3.62 | 3.70 | 4.00 | 5.46 | 16.7 |

Table 3: Preliminary statistics

| | Group | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SSE | 1-1 | 1-2 | 1-3 | 1-4 | 1-5 | 1-6 | 2-1 | 2-2 | 2-3 | 2-4 | 2-5 | 2-6 | 3-1 | 3-2 | 3-3 | 3-4 | 3-5 | 3-6 | 4-1 | 4-2 | 4-3 | 4-4 | 4-5 | 4-6 | Avg |
| Subjects min | 524 | 112 | 29.6 | 3.92 | 9.87 | 8.83 | 4.89 | 22.8 | 3.54 | 133 | 155 | 4.26 | 7.47 | 40.9 | 2.48 | 3.56 | 3.77 | 2.87 | 2.72 | 10.9 | 2.21 | 2.12 | 3.55 | 25.9 | - |
| Subjects max | 2354 | 1812 | 83.0 | 14.2 | 27.2 | 32.6 | 37.5 | 76.4 | 14.4 | 908 | 474 | 12.7 | 30.8 | 80.6 | 10.3 | 15.1 | 7.12 | 4.97 | 4.84 | 59.5 | 8.70 | 5.70 | 16.3 | 120 | - |
| Subjects mean | 1050 | 352 | 59.8 | 6.07 | 18.4 | 16.7 | 10.0 | 40.8 | 6.32 | 522 | 219 | 6.02 | 15.4 | 61.0 | 5.38 | 5.85 | 5.73 | 4.28 | 3.46 | 24.2 | 3.78 | 3.42 | 6.67 | 55.4 | - |
| Sticky info. (8) | 2110 | 1317 | 270 | 33.8 | 61.2 | 40.4 | 38.1 | 268.1 | 14.4 | 1720 | 1017 | 40.4 | 11.5 | 32.3 | 3.12 | 4.04 | 5.39 | 3.25 | 8.92 | 77.7 | 14.8 | 9.88 | 16.4 | 222 | 305 |
| Gen. mod. (18), $\zeta$=0 | 881 | 355 | 67.8 | 5.84 | 15.7 | 13.8 | 6.75 | 59.3 | 5.83 | 451 | 315 | 6.70 | 8.40 | 16.5 | 2.37 | 3.21 | 3.63 | 2.23 | 3.77 | 34.0 | 3.30 | 3.59 | 5.87 | 60.2 | 97 |
| Trend ext. (10) | 558 | 184 | 27.0 | 5.73 | 9.59 | 9.31 | 7.77 | 23.9 | 5.83 | 260 | 158 | 7.00 | 7.81 | 18.6 | 2.08 | 2.46 | 2.55 | 1.96 | 3.55 | 12.2 | 3.69 | 3.29 | 4.77 | 27.3 | 56 |
| General model (18) | 755 | 310 | 54.2 | 6.15 | 15.2 | 13.2 | 6.89 | 49.1 | 4.82 | 445 | 246 | 5.67 | 6.67 | 13.6 | 2.52 | 2.44 | 3.07 | 1.99 | 2.59 | 22.4 | 2.85 | 3.76 | 6.48 | $5.3e^8$ | 88 |
| Adaptive exp. (11) | 973 | 210 | 67.8 | 6.15 | 21.3 | 15.2 | 8.63 | 65.2 | 5.68 | 805 | 313 | 8.04 | 12.8 | 53.6 | 4.36 | 5.78 | 5.77 | 3.41 | 3.67 | 21.7 | 3.71 | 3.30 | 5.80 | 61.7 | 112 |

Table 4: Comparison between subjects and Classical Econometrician

| model (eq.) | Comparison 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Rational expectations: Stat (7) | 28.7 | 42.1 | - | - | - | - |
| Rational expectations: Theory (22) | - | - | 40.7 | 44.9 | - | - |
| AR(1) process (19) | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| Sticky information type (8) | 6.5 | 5.6 | 4.2 | 3.2 | 10.2 | 6.5 |
| Adaptive expectations (11) | 7.4 | 5.1 | 4.2 | 4.2 | 11.6 | 9.3 |
| Trend extrapolation (10) | 30.1 | 25.5 | 28.2 | 26.9 | 36.6 | 26.9 |
| Recursive - lagged inflation (13) | 11.6 | 7.9 | 8.8 | 8.3 | 21.8 | 9.3 |
| Recursive - REE (14) | 2.8 | 2.3 | 2.8 | 1.9 | 4.2 | 1.4 |
| Recursive - AR(1) process (15) | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | |
| Recursive - trend extrapolation (16) | 12.0 | 10.6 | 10.2 | 9.7 | 14.8 | 12.0 |
| General model (18), $\zeta = 0$ | - | - | - | - | - | 34.3 |

Table 5: Inflation expectation formation (percent of subjects)

| | Group | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model (eq.) | 1-1 | 1-2 | 1-3 | 1-4 | 1-5 | 1-6 | 2-1 | 2-2 | 2-3 | 2-4 | 2-5 | 2-6 | 3-1 | 3-2 | 3-3 | 3-4 | 3-5 | 3-6 | 4-1 | 4-2 | 4-3 | 4-4 | 4-5 | 4-6 | Avg |
| Gen. mod. (18), $\zeta$=0 | 17.7 | 18.5 | 24.6 | 16.2 | 14.8 | 11.4 | 13.3 | 14.5 | 14.3 | 18.5 | 19.5 | 16.8 | 12.1 | 16.3 | 27.9 | 18.4 | 10.3 | 26.4 | 11.4 | 15.7 | 13.6 | 22.2 | 12.8 | 19 | 16.9 |
| Sticky info. type (8) | 12.3 | 6.4 | 8.2 | 5.9 | 10.1 | 14.3 | 11.6 | 10.8 | 13.6 | 9.6 | 9.9 | 8.4 | 14.8 | 11.1 | 19.2 | 23.9 | 15.8 | 18.2 | 9.4 | 8.9 | 12.6 | 8.6 | 15.8 | 11.1 | 12.1 |
| Ad. exp. CGL (11) | 16 | 12.6 | 13 | 15.8 | 15 | 11.4 | 16.7 | 13.8 | 10.8 | 17.8 | 14.3 | 18.2 | 15.2 | 14.3 | 12.5 | 11.1 | 12.6 | 5.1 | 11.1 | 15.8 | 17.5 | 11.4 | 11.6 | 20.4 | 13.9 |
| Ad. exp. DGL (12) | 7.2 | 5.4 | 4.7 | 8.4 | 2.7 | 5.9 | 8.2 | 6.2 | 9.6 | 5.6 | 5.4 | 6.6 | 6.7 | 4.9 | 5.6 | 8.8 | 8.8 | 7.6 | 7.7 | 5.2 | 5.6 | 6.4 | 5.1 | 4 | 6.3 |
| Trend extr. (10) | 23.6 | 20 | 14.1 | 15 | 16.2 | 11.8 | 15.5 | 13.8 | 13.8 | 16.5 | 20.9 | 13.6 | 13.3 | 19.2 | 6.1 | 10.4 | 12.6 | 6.4 | 7.6 | 13.1 | 10.1 | 12.5 | 11.8 | 18.4 | 14 |
| Rec. AR(1) (15) | 7.9 | 12.1 | 11.1 | 19 | 19.2 | 12.8 | 19.2 | 9.4 | 20 | 8.4 | 8.8 | 18.9 | 11.1 | 10.8 | 12.5 | 13.5 | 22.2 | 18.7 | 26.1 | 16 | 29.1 | 19.7 | 19.7 | 7.1 | 15.6 |
| Rec. trend extr. (16) | 15.3 | 24.9 | 24.2 | 19.7 | 22.1 | 32.3 | 15.5 | 31.5 | 17.8 | 23.6 | 21.2 | 17.5 | 26.8 | 23.4 | 16.3 | 14 | 17.7 | 17.7 | 26.6 | 25.3 | 11.4 | 19.2 | 23.2 | 20 | 21.1 |
| inflation min | -9.53 | -5.27 | -0.84 | 0.67 | 0.81 | 1.16 | 1.26 | 0.06 | 1.84 | -9.04 | -6.74 | 0.8 | 2 | 0.97 | 2.49 | 2.41 | 2.48 | 2.51 | 2.4 | 1.77 | 1.88 | 2.46 | 1.77 | 0.68 | -9.53 |
| inflation max | 16.68 | 10.46 | 6.51 | 3.89 | 5.01 | 4.78 | 4.38 | 7.42 | 3.98 | 12.56 | 8.17 | 4.13 | 3.84 | 5.28 | 3.44 | 3.46 | 3.74 | 3.56 | 3.78 | 4.49 | 3.62 | 3.7 | 4 | 5.46 | 16.68 |
| inflation s.d. | 5.83 | 2.89 | 1.95 | 0.75 | 1.09 | 0.9 | 0.76 | 1.81 | 0.51 | 5.5 | 3.56 | 0.88 | 0.51 | 0.94 | 0.24 | 0.26 | 0.31 | 0.24 | 0.39 | 0.81 | 0.48 | 0.38 | 0.51 | 1.28 | 2.05 |

Table 6: Inflation expectation formation (percent of all cases)

| | Probit RE | Probit PA | Logit RE | Logit PA | Logit FE |
|---|---|---|---|---|---|
| Cons. | -0.2502*** | -0.2210*** | -0.4139*** | -0.3552*** | |
| | (0.0836) | (0.0749) | (0.1449) | (0.1188) | |
| $\lvert\pi_{t-1} - \pi_{t-2}\rvert$ | 0.0422 | 0.0402 | 0.0661 | 0.0639* | 0.0545 |
| | (0.0293) | (0.0247) | (0.0482) | (0.0388) | (0.0354) |
| $\pi_{t-1}$ | -0.0568*** | -0.0533*** | -0.0919*** | -0.0857*** | -0.076** |
| | (0.0219) | (0.0190) | (0.0345) | (0.0302) | (0.0383) |
| $y_{t-1}$ | -0.1702*** | -0.1596*** | -0.2747*** | -0.2577*** | -0.2540*** |
| | (0.0391) | (0.0381) | (0.0674) | (0.0623) | (0.0591) |
| $i_{t-1}$ | 0.0440** | 0.0415** | 0.0715** | 0.0670*** | 0.0575** |
| | (0.0181) | (0.0161) | (0.0286) | (0.0254) | (0.0275) |
| $\left(\pi_{t-1} - \pi^k_{t-1\mid t-2}\right)^2$ | 0.0061 | 0.006 | 0.011 | 0.0099 | 0.0089 |
| | (0.0171) | (0.0143) | (0.0248) | (0.0260) | (0.0359) |
| $\ln(\sigma^2)$ (panel) | -1.5874*** | | -0.5814*** | | |
| | (0.1996) | | (0.2064) | | |
| $\sigma$ (panel) | 0.4522*** | | 0.7478*** | | |
| | (0.0441) | | (0.0783) | | |
| $\rho$ (panel) | 0.1670*** | | 0.1453*** | | |
| | (0.0270) | | (0.0256) | | |
| N | 14040 | 14040 | 14040 | 14040 | 13975 |
| Groups | 216 | 216 | 216 | 216 | 215 |
| Obs per Group | 65 | 65 | 65 | 65 | 65 |
| Wald $\chi^2(9)$ | 34.0 | 31.8 | 31.2 | 32.6 | 36.2 |

Table 7: Determinants of swithing behavior. Notes: RE stands for random effects, PA population averages, while FE is for fixed effects model. Standard errors in parentheses. */**/*** denotes significance at 10/5/1 percent level. Standard errors are calculated using bootstrap procedures (1000 replications) that take into account potential presence of clusters in groups.

| Treatment | Groups | Comparison with Treatment 1 (p-value) |
|---|---|---|
| Inflation forc. targ. $\gamma = 1.5$ | $1 - 6$ | – |
| Inflation forc. targ. $\gamma = 1.35$ | $7 - 12$ | 0.6310 |
| Inflation forc. targ. $\gamma = 4$ | $13 - 18$ | 0.0104 |
| Inflation targeting $\gamma = 1.5$ | $19 - 24$ | 0.0250 |

Table 8: Comparison of standard deviations using Kruskal-Wallis test

| | Gen. mod. (18), $\zeta$=0 | | Sticky info. (8) | | ADE CGL (11) | | ADE DGL (12) | | Trend ext. (10) | | Rec. AR(1) (15) | | Rec. trend. extr. (16) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| s.e.: method | robust | cluster | robust | cluster | robust | cluster | robust | cluster | robust | cluster | robust | cluster | robust | cluster |
| $\rho_{js}$ | 4.6285 | 4.6285 | -12.2631*** | -12.2631* | 16.6381* | 16.6381* | -18.7886 | -18.7886 | 25.2361*** | 25.2361** | -15.6069*** | -15.6069 | 2.5229 | 2.5229 |
| | (-5.675) | (-7.324) | (-4.354) | (-4.569) | (-8.920) | (-6.612) | (-12.449) | (-17.302) | (-6.227) | (-7.832) | (-5.392) | (-7.622) | (-5.545) | (-6.258) |
| cons | 0.5816 | 0.5816 | 2.8509*** | 2.8509** | -0.9502 | -0.9502 | 2.5567*** | 2.5567 | -2.1698*** | -2.1698 | 3.7924*** | 3.7924* | 0.8321 | 0.8321 |
| | (-0.8462) | (-0.8648) | (-0.7462) | (-0.8232) | (-1.041) | (-0.6043) | (-0.8706) | (-1.268) | (-0.7605) | (-0.9535) | (-1.043) | (-1.504) | (-1.311) | (-1.586) |
| Observations | 24 | | 24 | | 24 | | 24 | | 24 | | 24 | | 24 | |
| $R^2$ | 0.0192 | | 0.1115 | | 0.1173 | | 0.0407 | | 0.4989 | | 0.3525 | | 0.0072 | |

Table 9: Relation of standard deviation to certain behavioral types as dened in Table 6. Notes: Standard errors in parentheses. */**/*** denotes signicance at 10/5/1 percent. Standard errors take into account potential presence of clusters in treatments.

| | reg1 | reg2 | reg3 | reg4 |
|---|---|---|---|---|
| $sd_{s,t}$ | 1.0147*** | 1.0121*** | 1.0121*** | 1.0099*** |
| | (0.0085) | (0.0073) | (0.0069) | (0.0066) |
| Gen. mod. (18), $\zeta = 0$ | 0.0018*** | 0.001 | 0.0031* | |
| | (0.0007) | (0.0013) | (0.0017) | |
| Sticky info. (8) | -0.0029* | -0.0039 | -0.0018 | -0.0043** |
| | (0.0016) | (0.0025) | (0.0019) | (0.0020) |
| ADE DGL (12) | -0.0023** | -0.0030** | -0.0008 | -0.0027** |
| | (0.0009) | (0.0013) | (0.0015) | (0.0014) |
| Trend Ext. (10) | 0.0067*** | 0.0055*** | 0.0077*** | 0.0055*** |
| | (0.0015) | (0.0018) | (0.0023) | (0.0014) |
| ADE CGL (11) | | -0.0011 | 0.001 | |
| | | (0.0018) | (0.0015) | |
| Recursive V1 (13) | | -0.0021 | | -0.0025 |
| | | (0.0025) | | (0.0018) |
| Recursive V4 (16) | | | 0.0021 | |
| | | | (0.0025) | |
| cons | -0.0759* | 0.0219 | -0.1895 | 0.0373 |
| | (0.0417) | (0.1378) | (0.1449) | (0.0556) |
| N | 1560 | 1560 | 1560 | 1560 |
| $\chi^2$ | 67328.4 | 54449.2 | 65883.1 | 79094.9 |

Table 10: Decision model's influence on standard deviation of inflation. Notes: Estimations are conducted using system GMM estimator of Blundell and Bond (1998) for dynamic panels. Standard errors in parentheses. */**/*** denotes significance at 10/5/1 percent level. Standard errors are calculated using bootstrap procedures (1000 replications) that take into account potential presence of clusters in treatments.

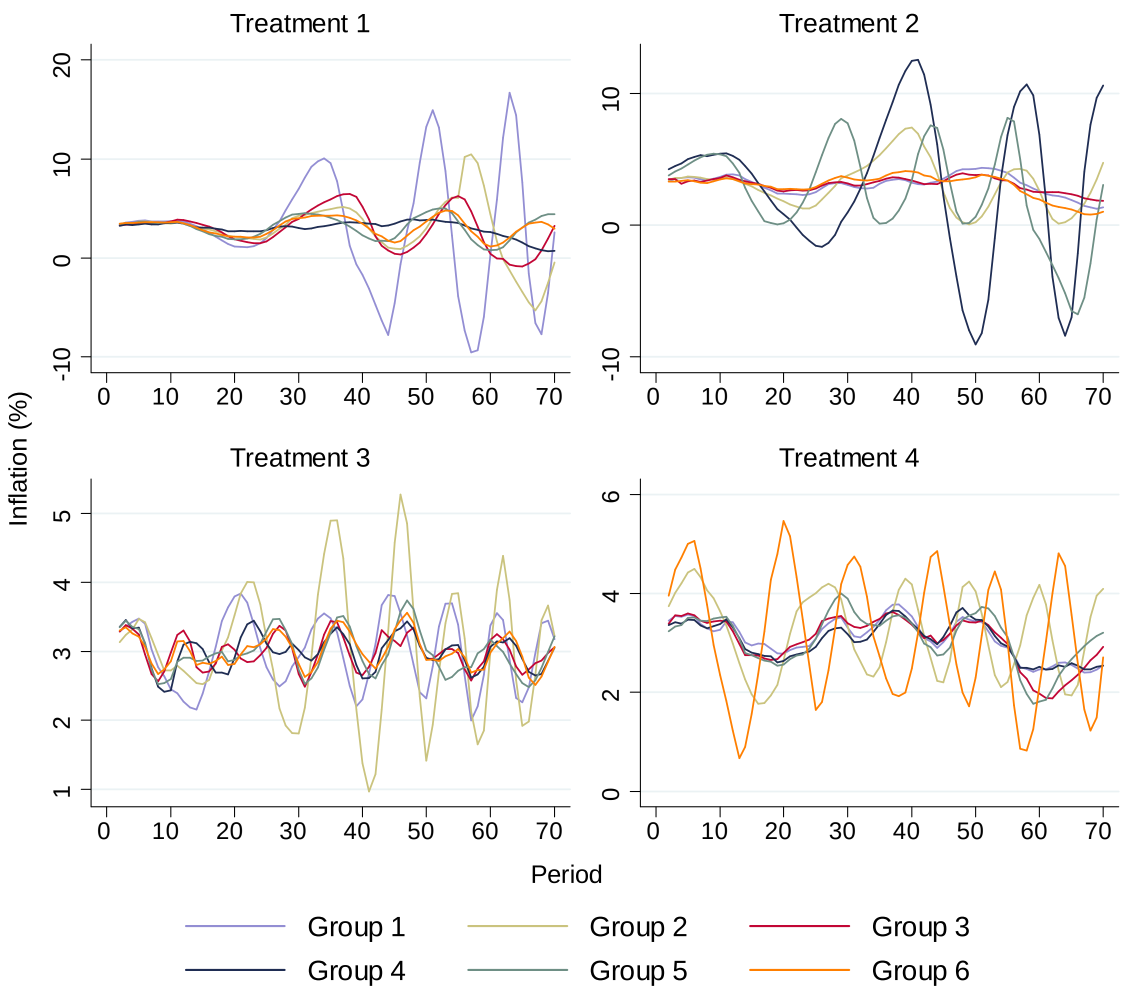


Figure 2: Group comparison of expected inflation (average subject prediction) and realized inflation by treatment. Treatment 1 has inflation forecast targeting (IFT) with $\gamma = 1.5$. Treatment 2 has IFT with $\gamma = 1.35$. Treatment 3 has IFT with $\gamma = 4$. Treatment 4 has inflation targeting with $\gamma = 1.5$.

# A Additional Tables and Figures

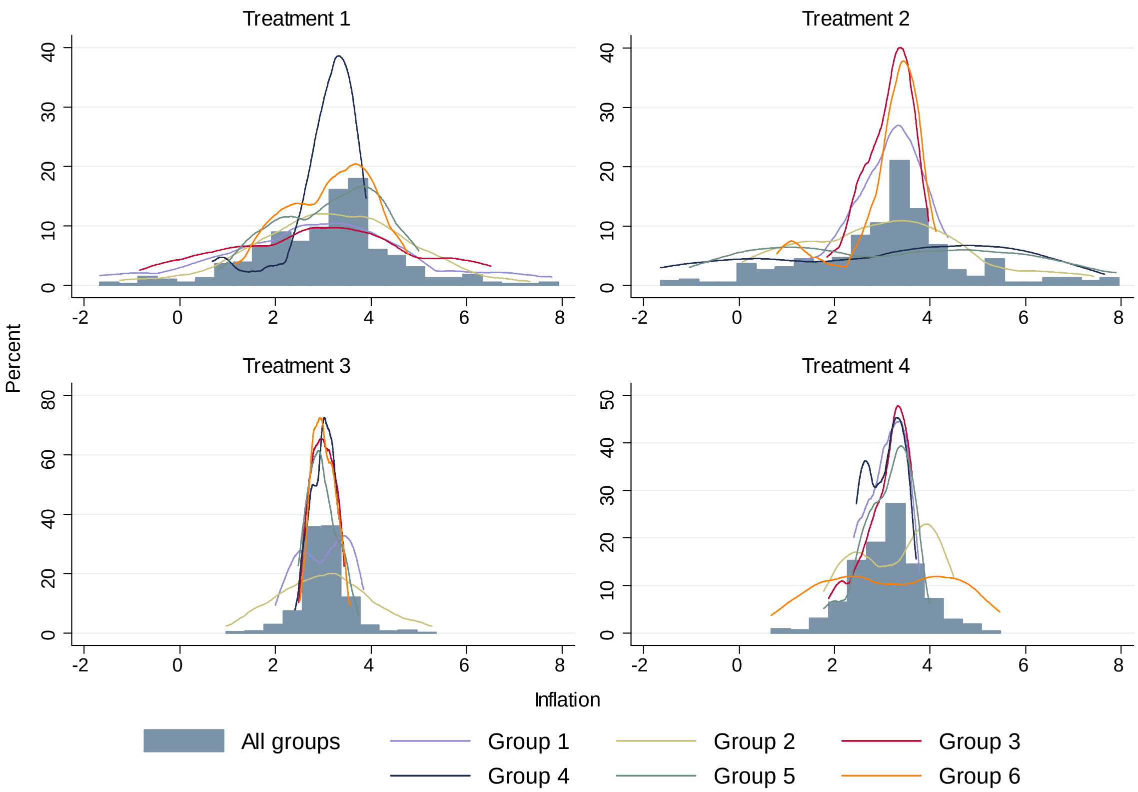


Figure A1: Histogram of individual inflation forecasts for the six independent groups in each treatment and combined

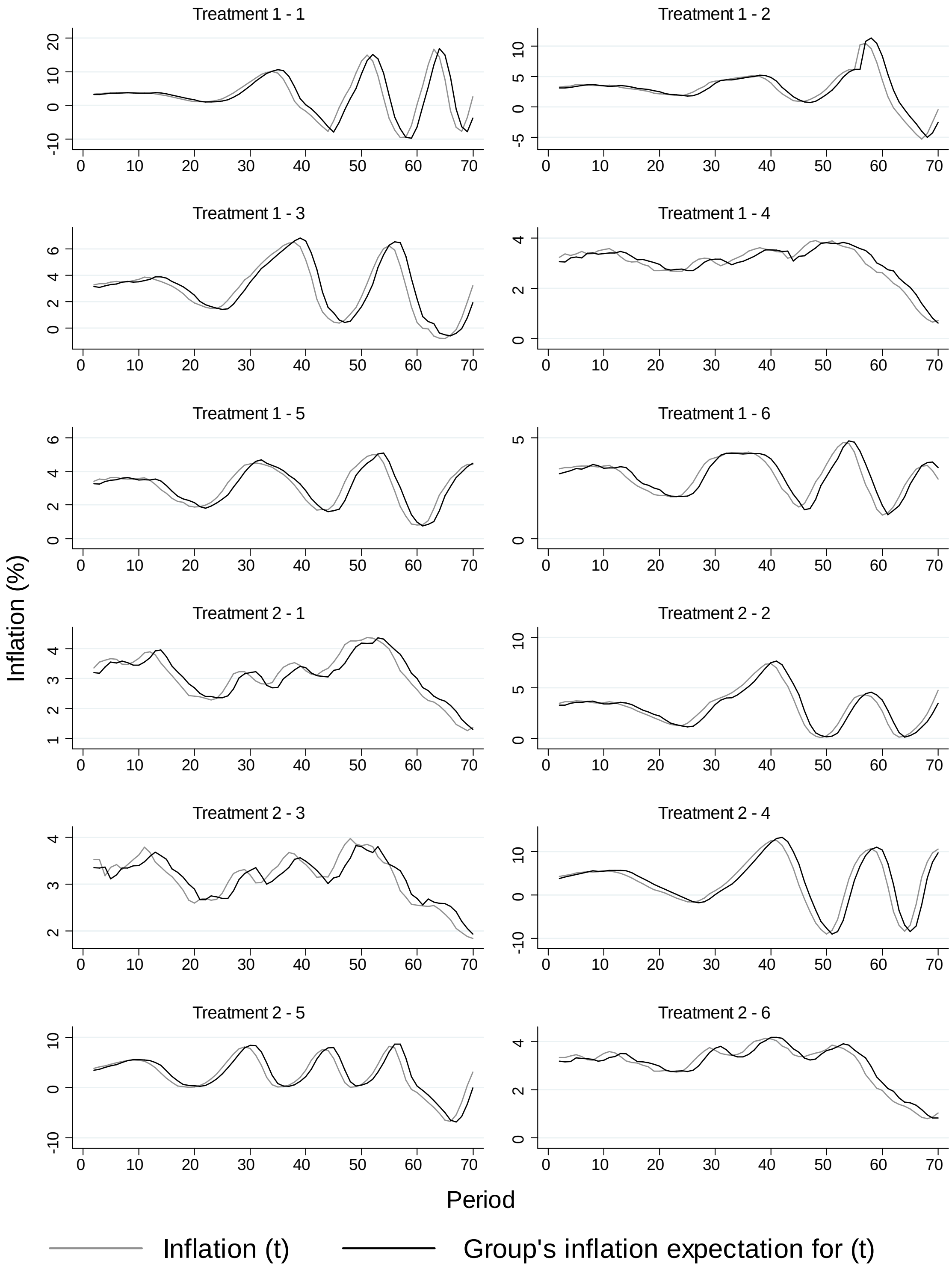


Figure A2: Inflation and inflation expectations per group, Part 1

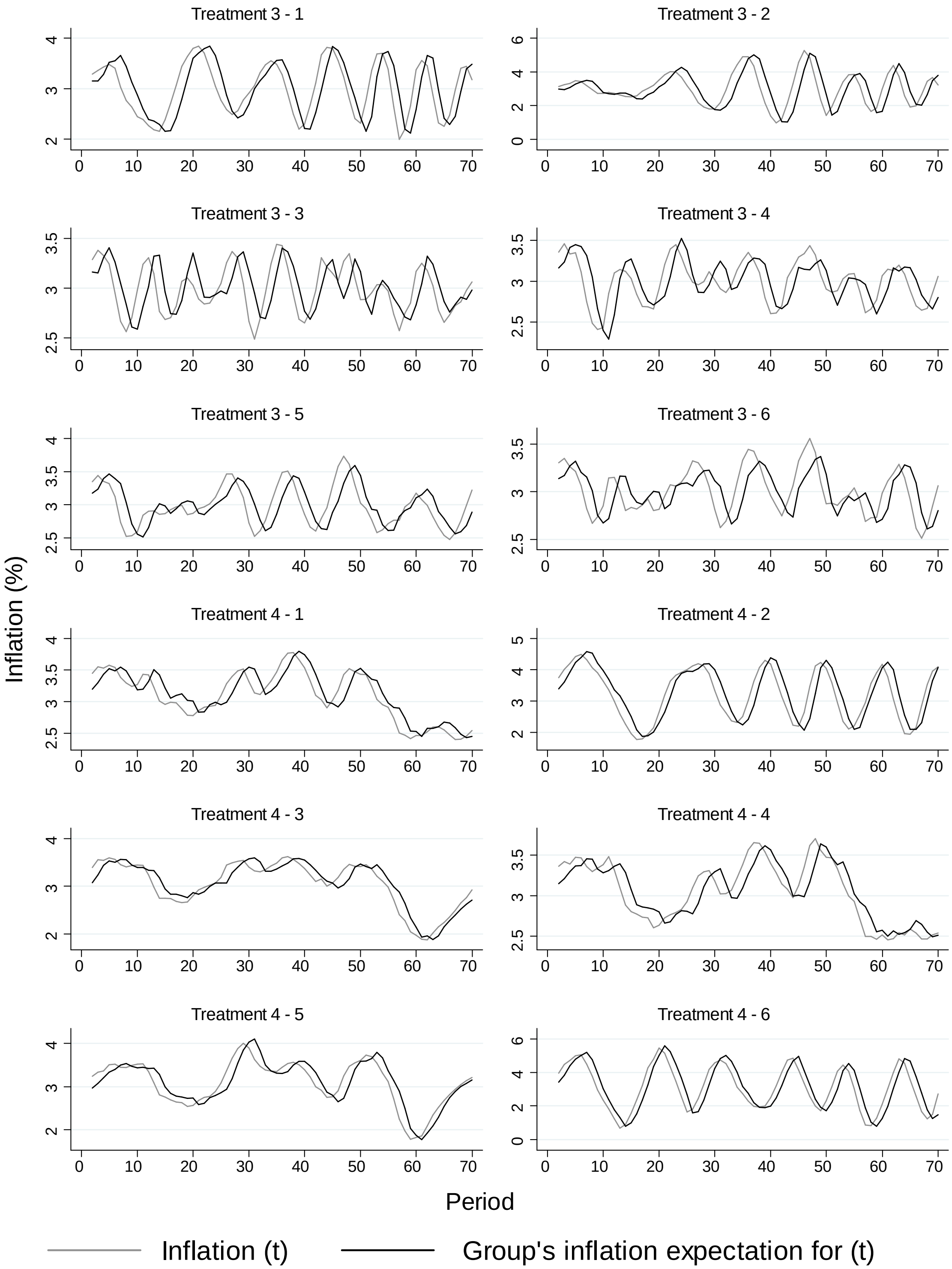


Figure A3: Inflation and inflation expectations per group, Part 2

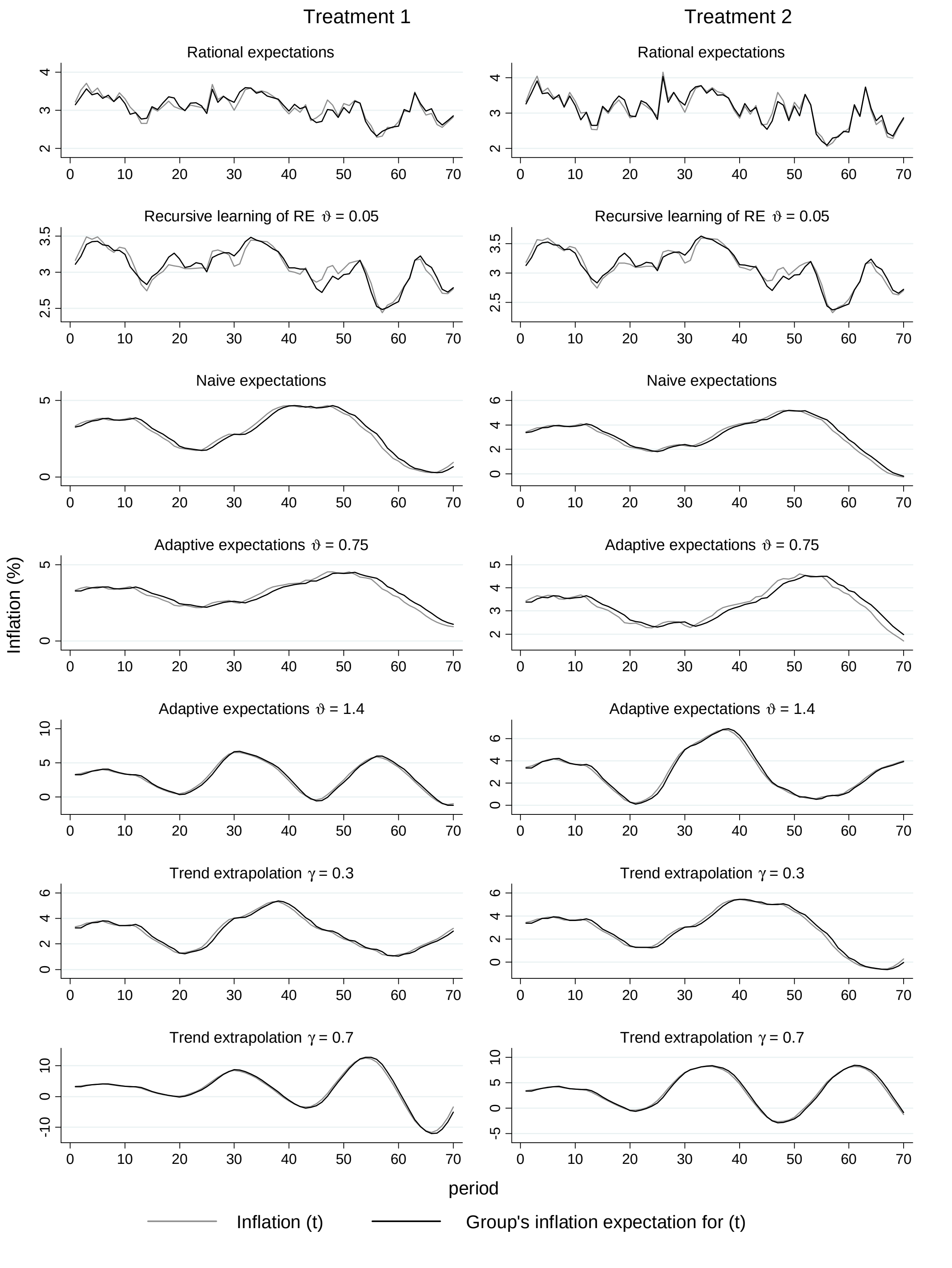


Figure A4: Alternative expectation formation rules (treatments 1 and 2).

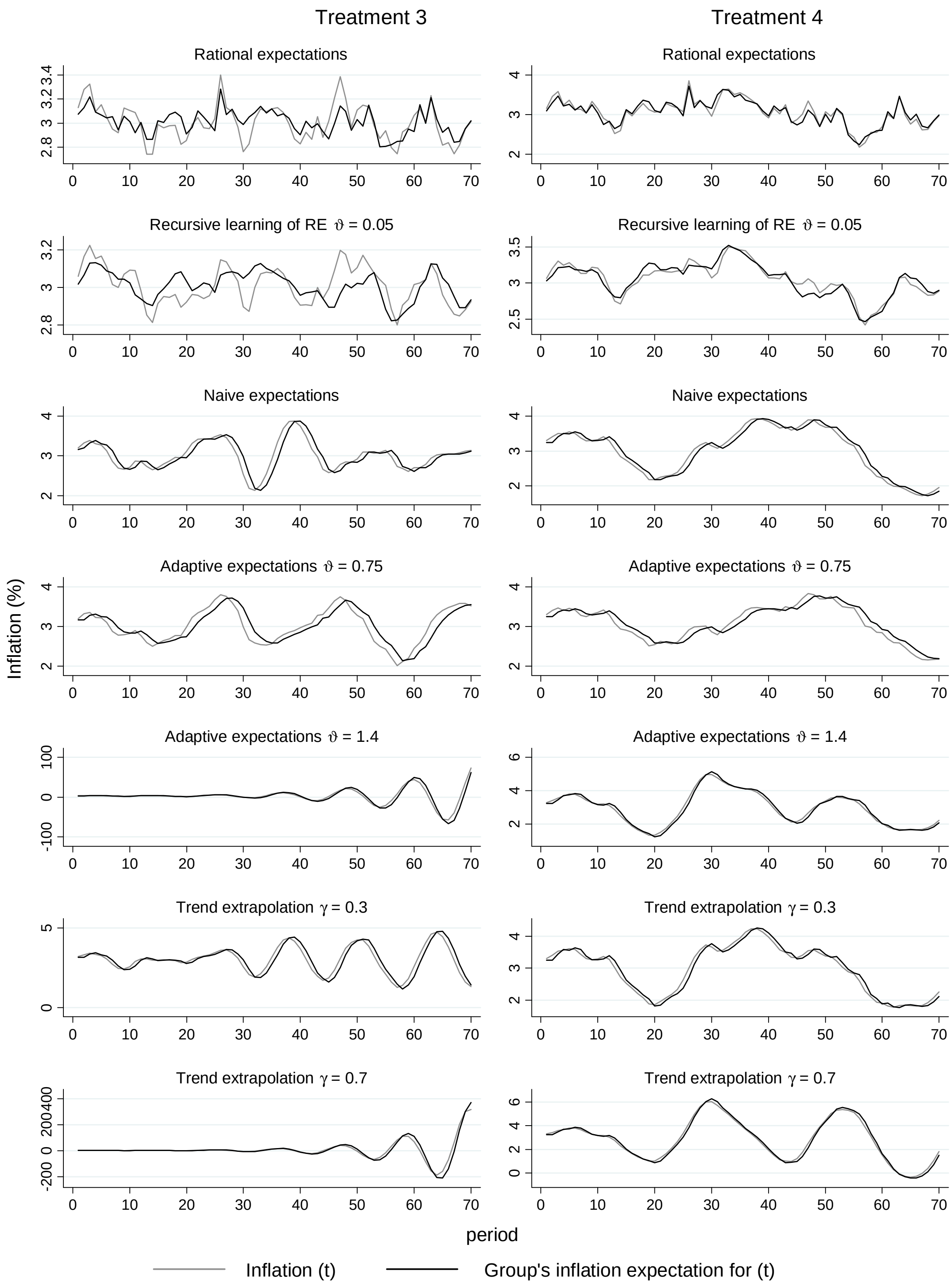


Figure A5: Alternative expectation formation rules (treatments 3 and 4).

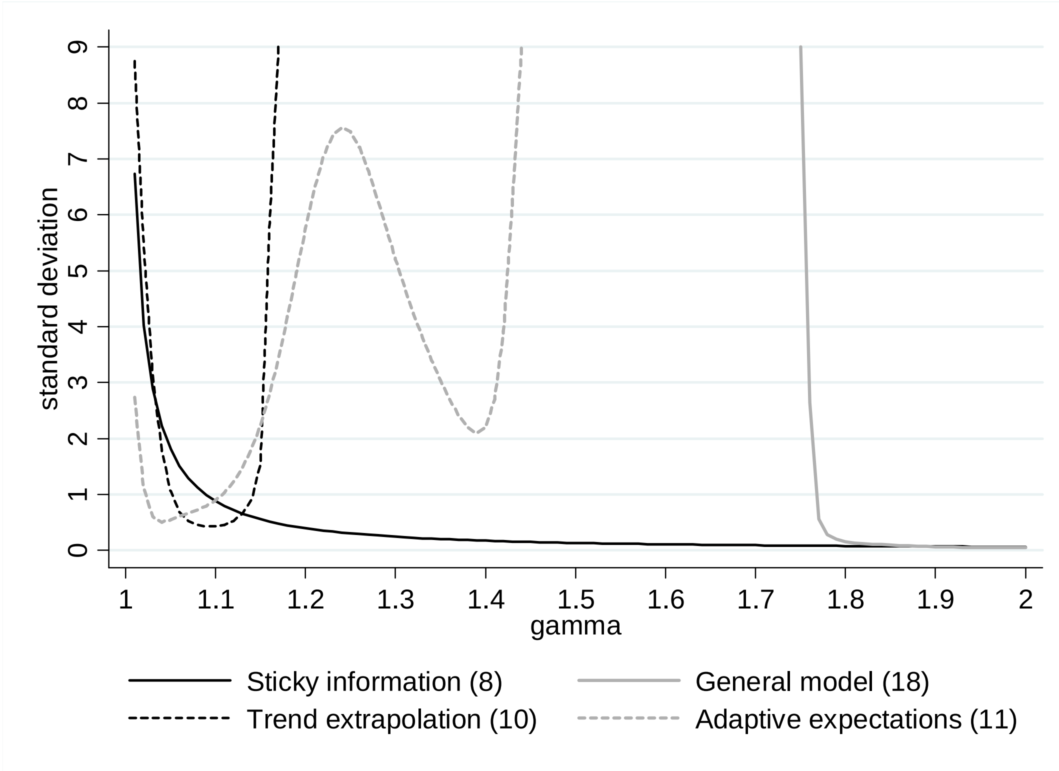


Figure A6: Variability of inflation and alternative expectation formation rules (inflation forecast targeting).

| model (eq.) | 1 | 2 | 3 | 4 | All |
|---|---|---|---|---|---|
| Rational expectations (7) | 35.2 | 48.1 | 5.6 | 25.9 | 28.7 |
| AR(1) process (19) | 0.0 | 0.0 | 0.0 | 1.9 | 0.5 |
| Sticky information type (8) | 3.7 | 1.9 | 16.7 | 3.7 | 6.5 |
| Adaptive expectations (11) | 9.3 | 3.7 | 7.4 | 9.3 | 7.4 |
| Trend extrapolation (10) | 35.2 | 25.9 | 25.9 | 33.3 | 30.1 |
| Recursive - lagged inflation (13) | 3.7 | 5.6 | 24.1 | 13.0 | 11.6 |
| Recursive - REE (14) | 0.0 | 1.9 | 9.3 | 0.0 | 2.8 |
| Recursive - trend extrapolation (16) | 0.0 | 0.0 | 0.0 | 1.9 | 0.5 |
| Recursive - AR(1) process (15) | 13.0 | 13.0 | 11.1 | 11.1 | 12.0 |

Table A1: Inflation expectation formation (percent of subjects, Comparison 1)

| | Probit RE | Probit PA | Logit RE | Logit PA | Logit FE |
|---|---|---|---|---|---|
| Cons. | -0.2502 | -0.221 | -0.4139 | -0.3552* | |
| | (0.3537) | (0.1884) | (0.3149) | (0.1817) | |
| $\lvert\pi_{t-1} - \pi_{t-2}\rvert$ | 0.0422 | 0.0402 | 0.0661 | 0.0639 | 0.0545 |
| | (0.0376) | (0.0317) | (0.0600) | (0.0559) | (0.0632) |
| $\pi_{t-1}$ | -0.0568 | -0.0533 | -0.0919 | -0.0857 | -0.076 |
| | (0.3717) | (0.1945) | (0.3099) | (0.1611) | (0.2747) |
| $y_{t-1}$ | -0.1702*** | -0.1596*** | -0.2747*** | -0.2577*** | -0.2540*** |
| | (0.0518) | (0.0471) | (0.0808) | (0.0769) | (0.0925) |
| $i_{t-1}$ | 0.044 | 0.0415 | 0.0715 | 0.067 | 0.0575 |
| | (0.2493) | (0.1308) | (0.2075) | (0.1072) | (0.1843) |
| $\left(\pi_{t-1} - \pi^k_{t-1\lvert t-2}\right)^2$ | 0.0061 | 0.006 | 0.011 | 0.0099 | 0.0089 |
| | (0.0432) | (0.0367) | (0.0688) | (0.0602) | (0.0518) |
| $\ln(\sigma^2)$ (panel) | -1.5874*** | | -0.5814*** | | |
| | (0.2882) | | (0.3014) | | |
| $\sigma$ (panel) | 0.4522*** | | 0.7478*** | | |
| | (0.0652) | | (0.1127) | | |
| $\rho$ (panel) | 0.1697*** | | 0.1453*** | | |
| | (0.0406) | | (0.0374) | | |
| N | 14040 | 14040 | 14040 | 14040 | 13975 |
| Groups | 216 | 216 | 216 | 216 | 215 |
| Obs per Group | 65 | 65 | 65 | 65 | 65 |
| Wald $\chi^2(9)$ | 26.2 | 38.5 | 25.1 | 25.1 | 27.1 |

Table A2: Determinants of swithing behavior. Notes: RE stands for random effects, PA population averages, while FE is for fixed effects model. Standard errors in parentheses. */**/*** denotes significance at 10/5/1 percent level. Standard errors are calculated using bootstrap procedures (1000 replications) that take into account potential presence of clusters in treatments.

# B (Not for Publication) Experimental Instructions[38]

Thank you for participating in this experiment, a project of economic investigation. Your earnings depend on your decisions and the decisions of the other participants. There is a show up fee of the 4 Euros assured. From now on until the end of the experiment you are not allowed to communicate with each other. If you have some question raise your hand and one of the instructors will answer the question in private. Please do not ask aloud.

[38] Instructions used for experiments at Universitat Pompeu Fabra are in Spanish language. In experimental sessions, they were accompanied with the screenshots of the experimental interface and the profit table with earnings for various combinations of *estimation error* and *confidence interval*.

**The Experiment**

All participants receive exactly the same instructions. You and 8 other subjects all participate as agents in the *same* fictitious economy. You will have to predict future values of given economic variables. The experiment consists of 70 periods. The rules are the same in all the periods. You will interact with the same 8 subjects during the whole experiment.

Imagine that you work in a firm where you have to predict inflation for the next period. Your profit depends on the accuracy of your inflation expectation.

**Information in Each Period**

The economy will be described with 3 variables in this experiment: the *inflation rate*, the *output gap*, and the *interest rate*.

- **Inflation** measures general rise in prices in the economy. Each period it depends on the inflation expectations of the agents in economy (you and other 8 participants in this experiment), output gap and small random shocks.

- The **output gap** measures for how much (in %) the actual Gross Domestic Product differs from the potential one. If the output gap is greater than 0, it means that the economy is producing more than the potential level, if negative, less than potential level. It depends each period on inflation expectations of the agents in economy, past output gap, interest rate and small random shocks.

- The **interest rate** is (in this experiment) the price of borrowing the money (in %) for one period. The interest rate is set by the monetary authority. Their decision mostly depends on inflation (expectations) of the agents in economy.

All given variables might be relevant for inflation forecast, but it is up to you to work out their relation and possible benefit of knowing them. The evolution of variables will partly depend on the inputs of you and other subjects and also different random shocks influencing the economy.

- You enter the economy in period 1. In this period you will be given computer generated past values of inflation, output gap and interest rate for 10 periods back (Called: -9, -8, … -1, 0)

- In period 2 you will be given all past values as seen in period 1 plus the value from period 1 (Periods: -9, -8, … 0, 1).

- In period 3 you will see all past values as in period 2 (Periods: -9, -8, ... 1, 2) plus YOUR prediction about inflation in period 2 that you made in period 1.

- In period $t$ you will see all past values of actual inflation up to period (Periods: -9, -8, ... , ) and your predictions up to period (Periods: 2, 3, ... , ).

**What Do You Have to Decide?**

Your payoff will depend on the accuracy of your prediction of the inflation in the future period. In each period your prediction will consist of two parts:

1. *Expected inflation*, (in %) that you expect to be in the NEXT period ($Exp.Inf.$)

2. The *Confidence Interval* ($Conf.Int.$) around your prediction for which you think there is 95% probability that the actual inflation will fall into. The interval is determined as the number of percentage points for which the actual inflation can be higher or lower.

**Example 1** *Let's say you think that inflation in the next period will be* 3.7%*. And you also think there is most likely (*95% *probability) that the actual inflation will not differ from that value for more than* 0.7 *percentage points. Therefore, you expect that there is* 95% *probability that actual inflation in the next period will be between* 3.0% *and* 4.4% *(*$3.7\% \pm 0.7\%$*). Your inputs in the experiment will be* 3.7 *under 1) and* 0.7 *under 2).*

Your goal is to maximize your payoff, given with the equation:

$$W = \max\left\{\frac{100}{1 + |Inflation - Exp.Inf.|} - 20, 0\right\} + \max\left\{\frac{100x}{1 + Conf.Int.} - 20, 0\right\}$$

where $Exp.Inf.$ is your expectation about the inflation in the NEXT period, $Conf.Int.$ is the confidence interval you have chosen, Inflation is the actual inflation in the next period, and $x$ is a variable with value 1 if

$$Exp.Inf. - Conf.Int. \leq Inflation \leq Exp.Inf. + Conf.Int.$$

and 0 otherwise.

This expression tells you, that $x$ will be 1, if actual inflation falls between $Exp.Inf. - Conf.Int.$ (3.0% in our example) and $Exp.Inf. + Conf.Int.$ (4.4% in our example).

The *first part* of the payoff function states that you will receive some payoff if the actual value in the next period will differ from your prediction in this period for less than 4 percentage points. The smaller this difference will be, the higher the payoff you receive. With a zero forecast error ($|Inflation - Exp.Inf.| = 0$), you would receive 80 units. However, if your forecast is 1 percentage point higher or lower than the actual inflation rate, you will get only 30 units ($100/2 - 20$). If your forecast error is 4 percentage points or more, you will receive 0 units ($100/5 - 20$).

The *second part* of the payoff function simply states that you will get some extra payoff if the actual inflation is within your expected interval and if that interval is not be larger than $\pm 4$ percentage point. The more certain of the actual value you are, the smaller interval you give, and the higher will be your payoff if the actual inflation indeed is in the given interval but there will also be higher chances that actual value will fall outside your interval. In our example this interval is $\pm 0.7$ percentage points. If the actual inflation falls in this interval you would receive 38.8 units ($100/(1 + 0.7) - 20$) in addition to the payoff from the first part of the payoff function. If the actual values is outside your interval, your receive 0.

In the attached sheet you can find table which shows various combinations of *forecast error* and *confidence interval* needed to earn a given number of points. See also figure on the next page.

**Information After Each Period**

Your payoff depends on your predictions for the next periods and actual realization in next period. Because the actual inflation will be only known in the next period, you will also be informed about you current period ($t$) prediction and earnings after the end of NEXT period ( $t + 1$). Therefore:

- After Period 1 you will not receive any earnings, since you did not make any prediction for the period 1.
- In any other period, you will receive the information about the actual inflation rate in this period and your inflation and confidence interval prediction from previous period. You will also be informed if the actual inflation value is in your expected interval and what are your earnings for this period.

The units in the experiment are fictitious. Your actual payoff will be the sum of profits from all the periods converted to euros in 1/500 conversion.

If you have any questions please ask them now!

**Questionnaire**[39]

1. If you believe that inflation in the next period will be _ _4.2%_ _, and you are quite sure that it will be higher than _ _ 3.5%_ _ and lower than _ _ 4.9%_ _, you will type:
   Under (1) _ _ _ _ _ _ _ _ _ _ for inflation, and
   Under (2) _ _ _ _ _ _ _ _ _ _ for confidence interval.

2. If you are now in period _ _ _ _ _ _ _ _ _ _, you have information about past inflation, output gap and interest rate up to period _ _ _ _ _ _ _ _ _ _ and you have to predict the inflation for period _ _ _ _ _ _ _ _ _ _.

[39]Options (1) and (2) are pointing to the different fields on the screenshot of the experimental interface.