TI 2015-087/II
Tinbergen Institute Discussion Paper

Monetary Policy under Behavioral Expectations: Theory and Experiment

Cars Hommes
Domenico Massaro
Matthias Weber

Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute, the Netherlands.

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

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

Tinbergen Institute has two locations:

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

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

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

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

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

Monetary Policy under Behavioral Expectations: Theory and Experiment*

Cars Hommes† Domenico Massaro† Matthias Weber‡

July 18, 2015

Abstract

Expectations play a crucial role in modern macroeconomic models. We replace the common assumption of rational expectations in a New Keynesian framework by the assumption that expectations are formed according to a heuristics switching model that has performed well in earlier work. We show how the economy behaves under these assumptions with a special focus on inflation volatility. Contrary to comparable models based on full rationality, the behavioral model predicts that inflation volatility can be lowered if the central bank reacts to the output gap in addition to inflation. We test the opposing theoretical predictions with a learning to forecast experiment. The experimental results support the behavioral model and the claim that reacting to the output gap in addition to inflation can indeed lower inflation volatility.

JEL classification: C90, E03, E52, D84

Keywords: Experimental Macroeconomics; Heterogeneous Expectations; Learning to Forecast Experiment; Trade-off Inflation and Output Gap

*Thanks for suggestions and comments go to Arthur Schram, participants of the North American ESA meetings in Fort Lauderdale, the UCSD-Rady Workshop on Incentives and Behavior Change in Modica, and participants of seminars in Amsterdam, Marseille, and Vilnius. Financial support from the EU 7th framework collaborative projects "Complexity Research Initiative for Systemic InstabilitieS (CRISIS)", grant no. 288501, and from The Netherlands' Organisation for Scientific Research (NWO), grant no. 406-11-022, are gratefully acknowledged.

†Center for Nonlinear Dynamics in Economics and Finance – CeNDEF, Amsterdam School of Economics (University of Amsterdam), and Tinbergen Institute

‡Center for Research in Experimental Economics and Political Decision Making – CREED, Amsterdam School of Economics (University of Amsterdam), and Tinbergen Institute

1 Introduction

Expectations play a crucial role in modern macroeconomic models that are used for scientific research and policy analysis. Usually, these expectations are modeled by assuming a representative fully rational agent. However, the assumption that all agents in an economy are fully rational and able to determine the model consistent expectation of the underlying process governing real-world economic outcomes is highly problematic. A lot of research has shown that humans are generally not able to react fully rationally to the world around them; this research ranges from providing evidence for simple but partly very persistent biases to showing the inability of humans to work with probabilities and to forecast future economic behavior (Tversky and Kahneman, 1974, Grether and Plott, 1979, Tversky and Thaler, 1990, and Kahneman et al., 1991, among many others; see Camerer et al., 2011 for a more recent overview). Also the claim based on evolutionary arguments that behavior deviating from the homogeneous rational expectations solution will be driven out of markets over time has been shown not to be generally true (Brock and Hommes, 1997, 1998, De Grauwe, 2012a; see also Arthur et al., 1997).

In this paper we replace the assumption of rational expectations in a macroeconomic model by the assumption that expectations are formed according to a behavioral heuristic switching model, which we take from earlier work. This particular behavioral model of expectation formation has been developed over a long period of time in which (mainly microeconomic) research has been conducted to investigate the questions of how people form expectations and of how they adapt their ways of forming expectations when confronted with observed economic outcomes (see Brock and Hommes, 1997, 1998, Carroll, 2003, Mankiw et al., 2003, Frankel and Froot, 1987, 1991, Bloomfield and Hales, 2002, Branch, 2004, Hommes, 2011, and Assenza et al., 2014b).

A key difference in outcomes between the macroeconomic models with rational and behavioral expectations concerns price stability, i.e. inflation volatility. Assuming rational expectations, there is a clear trade-off for a central bank between fighting inflation volatility and output gap volatility. If the central bank reacts to the output gap in addition to inflation this will under rational expectations always result in an increase of inflation volatility. This is different under behavioral expectations. Starting from a situation in which the central bank does not react to the output gap at all, the central bank can simultaneously decrease inflation volatility and output gap volatility by reacting to the output gap. However, inflation volatility as a function of the extent of output gap reaction is U-shaped. This means that reacting to the output gap on top of inflation will only lower inflation volatility up to a certain level after which inflation volatility starts to increase again.

These different outcomes regarding inflation volatility can be tested in the laboratory. We design a learning to forecast experiment where the only difference between treatments consists in the monetary policy rule used by the central bank. In one treatment the central bank only reacts to inflation, in the other it additionally reacts to the output gap. Our experimental results support the claim that inflation volatility can be lowered when the central bank also reacts to the output gap, in line with the predictions of the behavioral model.[1]

Our results from the behavioral model and the experimental data have clear policy implications for central banks with the sole aim of achieving price stability such as the European Central Bank.[2] Even if these banks only care about price stability, this goal is better achieved if they also react to changes in the output gap. This is important and at odds with standard macroeconomic thinking built upon full rationality.

This paper is organized as follows. In Section 2 we describe how we model the economy and the formation of expectations. We also show the main differences between the rational and behavioral versions. In Section 3 we first describe the experimental design and the procedures. Then we show the experimental results. Section 4 concludes.

2 Theory

In this section, we first describe the underlying macroeconomic model. Then we introduce the behavioral model of expectation formation. After that, we compare the outcomes of both models and describe the economic intuition behind these outcomes.

[1] This research builds upon various streams of literature; in particular on the literature on experimental macroeconomics and learning to forecast experiments (e.g. Marimon and Sunder, 1993, Van Huyck et al., 1994, Bernasconi and Kirchkamp, 2000, Kelley and Friedman, 2002, Lei and Noussair, 2002, Arifovic and Sargent, 2003, Hommes et al., 2005b, Adam, 2007, Heemeijer et al., 2009, Davis and Korenok, 2011, Bao et al., 2012, Kryvtsov and Petersen, 2013, Cornand and M'Baye, 2013, Pfajfar and Zakelj, 2014, Assenza et al., 2014b; see Duffy, 2012, and Assenza et al., 2014a, for surveys) and on the literature on behavioral macroeconomics (in particular works that consider monetary and fiscal policy when allowing for a departure from the hypothesis of rational expectations; e.g. Bullard and Mitra (2002), Marcet and Nicolini (2003), Guesnerie (2009), Branch and McGough (2009, 2010), Woodford (2010), De Grauwe (2011, 2012a,b), De Grauwe and Kaltwasser (2012), Anufriev et al. (2013), Kurz et al. (2013), Benhabib et al. (2014); see Evans and Honkapohja (2001) and Woodford (2013) for overviews).

[2] There are many other central banks with a hierarchical mandate which makes price stability the primary objective for monetary policy, including the central banks of New Zealand, Canada, England, and Sweden.

2.1 Macroeconomic Model

The economic model we use can be described by the following aggregate New Keynesian equations:

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

$$\pi_t = \lambda y_t + \rho \bar{\pi}^e_{t+1} + u_t \tag{2}$$

$$i_t = \text{Max}\{\bar{\pi} + \phi_\pi(\pi_t - \bar{\pi}) + \phi_y(y_t - \bar{y}),\, 0\}, \tag{3}$$

where y_t and $\bar{y}^e_{t+1}$ are respectively the actual and average expected output gap, i_t is the nominal interest rate, π_t and $\bar{\pi}^e_{t+1}$ are respectively the actual and average expected inflation rates, g_t and u_t are exogenous disturbances and φ, λ, ρ, ϕ_π and ϕ_y are positive parameters. Equation (1) is the aggregate demand equation in which the output gap y_t depends on the average expected future output gap $\bar{y}^e_{t+1}$ and on the real interest rate $i_t - \bar{\pi}^e_{t+1}$. Equation (2) is the New Keynesian Phillips curve according to which the inflation rate depends on the output gap and on average expected future inflation. Equation (3) is the monetary policy rule implemented by the central bank describing how it reacts to deviations from the inflation target $\bar{\pi}$ and to deviations from the corresponding equilibrium level of the output gap $\bar{y} \equiv (1-\rho)\bar{\pi}/\lambda$. The coefficients ϕ_π and ϕ_y measure how much the central bank adjusts the nominal interest rate i_t in response to deviations of the inflation rate from its target and of the output gap from its equilibrium level. As usual, the interest rate rule is subject to the zero lower bound, i.e. $i_t \geq 0$. When the zero lower bound is not binding, model (1)–(3) can be rewritten in matrix form as

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \Omega \begin{bmatrix} \varphi\bar{\pi}(\phi_\pi - 1) + \varphi\phi_y\bar{y} \\ \lambda\varphi\bar{\pi}(\phi_\pi - 1) + \lambda\varphi\phi_y\bar{y} \end{bmatrix} + \Omega \begin{bmatrix} 1 & \varphi(1-\phi_\pi\rho) \\ \lambda & \lambda\varphi + \rho + \rho\varphi\phi_y \end{bmatrix} \begin{bmatrix} \bar{y}^e_{t+1} \\ \bar{\pi}^e_{t+1} \end{bmatrix} + \Omega \begin{bmatrix} 1 & -\varphi\phi_\pi \\ \lambda & 1+\varphi\phi_y \end{bmatrix} \begin{bmatrix} g_t \\ u_t \end{bmatrix}, \tag{4}$$

where $\Omega \equiv 1/(1 + \lambda\varphi\phi_\pi + \varphi\phi_y)$.

The economic model described by the aggregate equations (1)–(3), or equivalently by (4), has microfoundations under both rational expectations and under behavioral expectations.[3] In the following we will only make use of the aggregate equations

[3]Under the assumption of a representative agent holding rational expectations, this model represents the standard New Keynesian model discussed for example in Woodford (2003) and Galí (2008). Micro-founded New Keynesian models consistent with heterogeneous expectations have been derived by Branch and McGough (2009), Kurz (2011), Kurz et al. (2013) and Massaro (2013). System (1)–(3) corresponds to the model developed by Branch and McGough (2009) augmented with demand and supply shocks, or to the model derived in Kurz (2011) and Kurz et al. (2013) in which the error terms are interpreted as the deviation of the average of agents' forecasts of their individual future consumption from the average forecast of aggregate consumption and as a similar deviation of price

presented here.

2.2 A Behavioral Model of Expectation Formation

Models with rational expectations are based on the assumption that agents have perfect information and a full understanding of the true model underlying the economy. There is, however, a large body of empirical literature documenting departures from this assumption and showing that agents use heuristics to make forecasts of future (macroeconomic) variables; this behavior is not necessarily a consequence of agents' irrationality but it can also be a "rational" response of agents who face cognitive limitations and have imperfect understanding of the true model underlying the economy (see e.g. Gigerenzer and Todd, 1999, or Gigerenzer and Selten, 2002). Next, we introduce a behavioral model of expectation formation for such an environment.

Let $\mathcal{H}$ denote a set of H different heuristics used by agents to make forecasts. A generic forecasting heuristic $h \in \mathcal{H}$ based on available information at time t can be described as

$$x^e_{h,t+1} = f_h(x_{t-1}, x_{t-2} \ldots; x^e_{h,t}, x^e_{h,t-1} \ldots). \tag{5}$$

In this paper x is either inflation π or the output gap y. Although agents might use simple rules to predict future inflation and output gap, we impose a certain discipline in the selection of such rules in order to avoid completely irrational behavior. In particular, we introduce a selection mechanism that disciplines the choice of heuristics by agents according to a fitness criterion. This allows agents to learn from past mistakes (the willingness to learn from past mistakes has been called "the most fundamental definition of rational behaviour"; De Grauwe, 2012b). We denote by U_h the fitness measure of a certain forecasting strategy h defined as

$$U_{h,t-1} = F(x^e_{h,t-1} - x_{t-1}) + \eta U_{h,t-2} \,, \tag{6}$$

where F is a generic function of the forecast error of heuristic h, and $0 \leq \eta \leq 1$ is a memory parameter, measuring the relative weight agents give to past errors of heuristic h. Performance is completely determined by the most recent forecasting error if $\eta = 0$, while performance depends on all past prediction errors with exponentially declining weights if $0 < \eta < 1$ or with equal weights if $\eta = 1$. If all agents simultaneously update the forecasting rule they use, the fraction of agents choosing rule h in

forecasts.

each period t can be described as

$$n_{h,t} = \frac{\exp\left(\beta U_{h,t-1}\right)}{\sum_{h=1}^{H} \exp\left(\beta U_{h,t-1}\right)} \,. \tag{7}$$

The multinomial logit expression described in Equation (7) can be derived directly from a random utility model (see Manski and McFadden, 1981, and Brock and Hommes, 1997). The parameter $\beta \geq 0$, referred to as "intensity of choice", reflects the sensitivity of agents to selecting the optimal prediction strategy according to the fitness measure U_h.[4] If $\beta = 0$, $n_{h,t}$ is constant for all h, meaning that agents do not exhibit any willingness to learn from past performance; if $\beta = \infty$ all agents adopt the best performing heuristic with probability one. The reinforcement learning model in Equation (7) has been extended by Hommes et al. (2005a) and Diks and van der Weide (2005) to include asynchronous updating in order to allow for the possibility that not all agents update their rule in every period (consistent with empirical evidence; see Hommes et al., 2005b, and Anufriev and Hommes, 2012). This yields a generalized version of Equation (7) described by

$$n_{h,t} = \delta n_{h,t-1} + (1-\delta)\frac{\exp\left(\beta U_{h,t-1}\right)}{\sum_{h=1}^{H} \exp\left(\beta U_{h,t-1}\right)} \,. \tag{8}$$

The parameter $0 \leq \delta \leq 1$ introduces persistence in the adoption of forecasting strategies and can be interpreted as the average fraction of individuals who, in each period, stick to their previous strategy.

In order to use this behavioral model for policy analyses or predictions, specific assumptions have to be made on the nature of agents' forecasting heuristics (in general, the set $\mathcal{H}$ may contain an arbitrary number of forecasting rules). We restrict our attention to a set of four heuristics described in Table 1.

The choice of this specific set of heuristics is motivated on empirical grounds. These heuristics were obtained and estimated as descriptions of typical individual forecasting behavior observed in Hommes et al. (2005b), Hommes et al. (2008), and Assenza et al. (2014b) building upon a rich literature on expectation formation (see Hommes, 2011, for a recent survey).[5] Based upon the calibration in these papers, we use the

[4]Equation (7) can also be derived from an optimisation problem under rational inattention (see Matějka and McKay, 2015). In this context the parameter β is inversely related to the "shadow cost of information".

[5]The reinforcement learning model described in Equation (8) including the set of heuristics presented in Table 1 is successfully used in Anufriev and Hommes (2012) to explain different price patterns observed in asset pricing experiments. In Assenza et al. (2014b) the model is used to explain the

Table 1: Set of heuristics

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

Notes: x^{av}_{t-1} denotes the average of all observations up to time $t-1$.

parameters $\beta = 0.4$, $\delta = 0.9$, and $\eta = 0.7$.[6]

2.3 Monetary Policy, Inflation, and Output Gap

A result derived from Model (4) under rational expectations is that a policy trade-off is observed between the volatility of the output gap and the volatility of inflation. A decline in output gap volatility resulting from a more active output stabilization policy comes at the price of an increase in inflation volatility (it is reasonable to focus on volatility as for the rational and the behavioral model alike inflation and output gap are on average at their target, respectively steady state level, for reasonable values of ϕ_π and ϕ_y). This policy trade-off is described in Figure 1a, where we show the effect of the parameter ϕ_y (with which the central bank reacts to deviations of the output gap from its steady state level) on inflation volatility. Higher output stabilization, i.e., an increase in the reaction coefficient ϕ_y, comes at the price of higher inflation volatility. The immediate policy implication for a central bank whose main objective is price stability is that it is optimal to set $\phi_y = 0$, i.e. not to react to output gap fluctuations at all (cf. Galí, 2008, and Woodford, 2003).

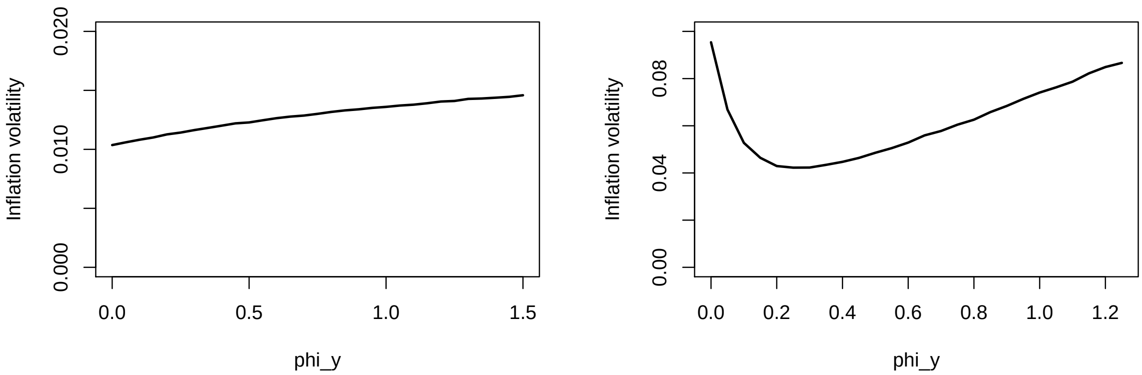
For the simulations of this graph, the parameter ϕ_π is equal to 1.5 (using different parameters of ϕ_π leads to similar results, which can be seen in Appendix A) and the structural parameters in Equations (1)-(3) are as estimated in Clarida et al. (2000).[7] The inflation target used for the simulations is $\bar{\pi} = 3.5$ (this is the same target that

observed patterns of inflation and output gap in a learning to forecast experiment framed in a New Keynesian model similar to (4).

[6]Furthermore, we use the fitness measure $U_h = 100/(1 + |x^e_h - x|)$, which is the function used to incentivize subjects in the experiment described in Section 3 (this incentive structure is also used in Adam, 2007, Pfajfar and Zakelj, 2014, and Assenza et al., 2014b, among others). The simulation results in Section 2.3 are qualitatively robust to alternative specifications of the fitness metric, such as using a quadratic function.

[7]Thus, $\rho = 0.99$, $\lambda = 0.3$, and $\varphi = 1$ (for quarterly data). The shocks g_t and u_t are independent and normally distributed with standard deviation 0.1. The number of simulations for each value of ϕ_y is 10000.

will be used in the experiment, Footnote 12 provides a rationale for this value; the simulations yield similar results for different values of $\bar{\pi}$). This inflation target leads to a steady state level of the output gap of $\bar{y} = 0.1166667$. Inflation volatility is measured by $v(\pi) = \frac{1}{T-1}\sum_{t=2}^{T}(\pi_t - \pi_{t-1})^2$, with T denoting the total number of periods. This measure has some properties that make it preferable to other measures of volatility (using alternative measures yields similar results).[8]



(a) Rational model

(b) Behavioral model

Figure 1: Inflation volatility as a function of ϕ_y for the rational and the behavioral model

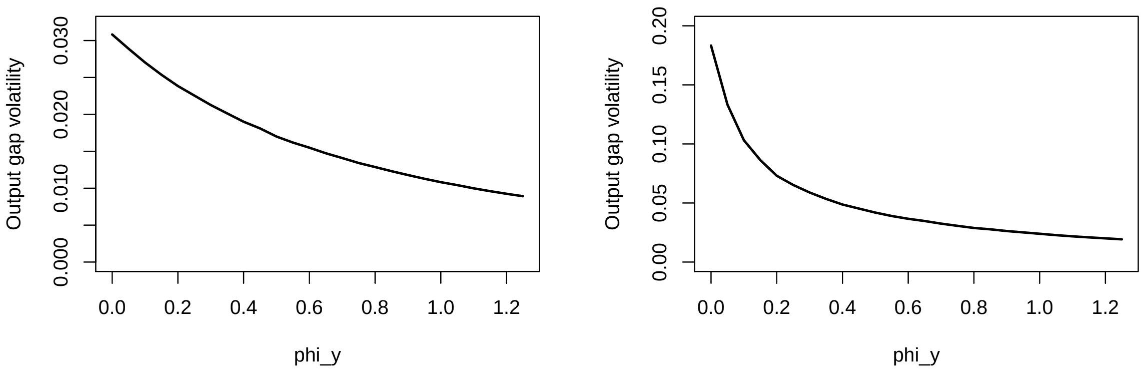
Notes: This figure shows the effect of parameter ϕ_y on inflation volatility. $\phi_\pi = 1.5$ for both sub-figures.

In Figure 1b, we show the effect of the parameter ϕ_y on inflation volatility when expectations are formed according to the behavioral model described in Section 2.2 (note that the scales in Figures 1a and 1b are different; the overall level of inflation volatility is higher under behavioral expectations than under rational expecations). In contrast to the simulation results under rational expectations, the graph of inflation volatility as a function of ϕ_y has a U-shape.[9] Thus, starting from $\phi_y = 0$, the central bank can simultaneously decrease volatility of inflation and output gap by also reacting with its monetary policy to deviations of the output gap from its steady state level (Figure 2 depicts output gap volatility as a function of ϕ_y; as ϕ_y increases output gap

[8] One could also use $v(\pi) = \frac{1}{T-1}\sum_{t=2}^{T}|\pi_t - \pi_{t-1}|$ or simply the standard deviation to measure volatility. An advantage of the measure we use when compared to the standard deviation is that short-term fluctuations are accounted for differently. The standard deviation of a time series does not change after a permutation of its values, although this can change how much the series fluctuates (imagine one time series that always alternates between the same level of high and low and one that first stays at the same low value for a while and then switches to the same high value).

[9] The starting values used for the simulations of the behavioral model are $\pi_{start} = 3.0$ and $y_{start} = 0.5$, Appendix A provides graphs for different starting values, which are also U-shaped.

volatility decreases under both rational and behavioral expectations). These results are qualitatively robust to changes in starting values, inflation target, and parameters involved in the expectation formation.[10] Hence, under behavioral expectations, there is a broader scope for output stabilization.



(a) Rational model (b) Behavioral model

Figure 2: Output gap volatility as a function of ϕ_y for the rational and the behavioral model

Notes: This figure shows the effect of parameter ϕ_y on output gap volatility. $\phi_\pi = 1.5$ for both subfigures.

Now we turn to the intuition of these results. Considering the outcome simulated with rational expectations (Figure 1a), it would be easy to fall prey to the following simple, but incorrect, intuition: "If there are two variables, targeting one variable will always come at the expense of the other variable". This is in general not the case, the intuition is slightly more complex. Homogeneous rational expectations are strictly forward looking and in this model always equal to the inflation target and the corresponding steady state level of the output gap, respectively (assuming that $\phi_\pi + \phi_y(1-\rho)/\lambda > 1$, which ensures a determinate model solution, see e.g. Woodford, 2003). These expectations do not depend in any way on the current level of inflation and output gap or on any past behavior. It is exactly via the dependence of expectations on (past) actual variables that reacting to the output gap can also pay off in terms of inflation volatility. To illustrate this, imagine inflation and output gap staying constant at $\bar{\pi}$ and $\bar{y}$ and a combination of shocks arriving in one period that would lead (without any

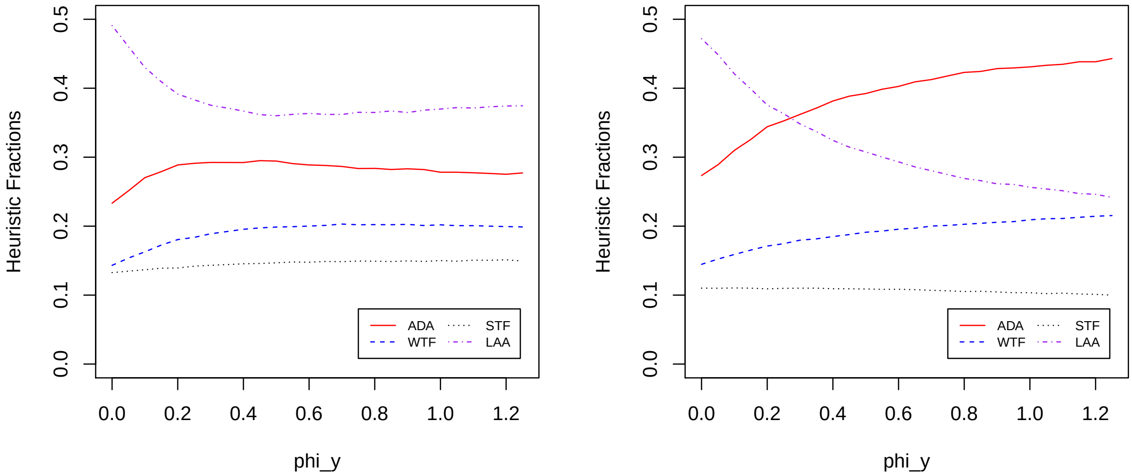
[10]Similar results also arise in a different macroeconomic model when employing simplistic behavioral rules of expectation formation (De Grauwe, 2011, 2012a). Such a non-monotonic trade-off between inflation and output gap volatility can also arise in sticky information economies in which the degree of attentiveness or the rate at which agents update their information is endogenized (Branch et al., 2009).

reaction by the central bank) to inflation staying constant and the output gap being above the steady state level. Should the central bank react to this shock if it only cares about inflation? The rational expectations answer would be no; inflation is at its target and in the next period one would (assuming no further shocks) again be at the inflation target and the steady state level of the output gap, because expectations do not react to the past. However, under behavioral expectations, what happens today matters for the future. If there is some trend-following behavior, a higher output gap now will lead agents to revise their expectations of the future output gap upwards, which will in turn lead to a higher realized output gap in the future which will lead to upward pressure on inflation. Therefore it can be beneficial for the central bank to curb the increase of the output gap now (at the expense of slightly lower inflation now) in order to dampen the upward pressure on inflation in the future. However, if the monetary authority puts too much weight on output gap stabilization, the ensuing fluctuations in inflation dominate the stabilization bonus provided by less volatile output, leading to higher inflation volatility.

Regarding the results from the behavioral model, one can also look at the heuristics involved. Of the four heuristics, only one heuristic is stabilizing, namely the adaptive rule (ADA). All other heuristics have some component extrapolating trends (coordination on trend-following heuristics is generally associated with high volatility, see e.g. Anufriev and Hommes, 2012, Bao et al., 2012, Assenza et al., 2014b, and Pfajfar and Zakelj, 2014). Thus, the volatility of a variable tends to be lower when the adaptive rule performs relatively well. Therefore, values of ϕ_y that lead to a relatively large fraction of the adaptive rule for inflation forecasting can be expected to also lead to low inflation volatility. This is indeed the case and can be seen in Figure 3, where the average fractions of heuristics used for inflation forecasting and output gap forecasting are shown as a function of ϕ_y. For inflation forecasting, starting from $\phi_y = 0$, increasing the reaction to the output gap first increases the fraction of agents using the adaptive rule, but after some level of ϕ_y, increasing it further reduces the fraction of agents using the adaptive rule. For output gap forecasting, it is not surprising to see that the fraction of agents using the adaptive rule increases monotonically with ϕ_y.

3 Experiment

The only task for subjects in the experiment is to forecast inflation and output gap. These forecasts are then used to calculate subsequent realizations. The model under-



(a) Heuristics Inflation

(b) Heuristics Output Gap

Figure 3: Fractions of heuristics used for behavioral expectations of output gap and inflation

Notes: This figure shows the average fractions of heuristics used as functions of ϕ_y ($\phi_\pi = 1.5$).

lying the experimental economy is the macroeconomic model described in Section 2.1 (with the same calibration of macroeconomic parameters as before). Before we describe the experiment in more detail, we now explain the treatments and hypotheses. The design of the experiment and the hypotheses can be motivated with the theory described in Section 2; however, the experiment is also informative without this theory in the background, as it can be seen as a mere investigation of the effects of a change in monetary policy in a controlled laboratory environment.

3.1 Treatments and Hypotheses

There are two treatments, $T1$ ("inflation targeting only") and $T2$ ("inflation and output gap targeting"). The only difference between the treatments lies in the Taylor rule describing monetary policy. In $T1$, the parameters of the Taylor rule are $\phi_\pi = 1.5$ and $\phi_y = 0$, whereas they are $\phi_\pi = 1.5$ and $\phi_y = 0.5$ in $T2$. That is, the only difference between the treatments is that in $T1$ the central bank only targets inflation whereas it also targets the output gap on top of inflation in $T2$.

We are interested in testing the null-hypothesis (which can be derived from the rational expectations model in Section 2) that inflation volatility in $T1$ is less or equal to inflation volatility in $T2$ against the alternative hypothesis (which can be derived from the behavioral model) that inflation volatility is greater in $T1$ than in $T2$. Figure 4 summarizes these hypotheses, i.e. the treatment effects one can expect arising from rational expectations and from the behavioral model described.[11]

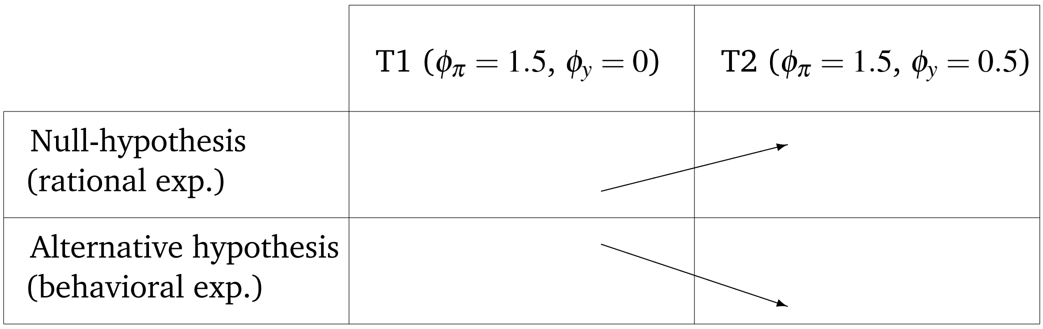


Figure 4: Hypotheses about inflation volatility

The inflation and output gap expectations arising from a continuum of rational agents are $\bar{\pi}^e = \bar{\pi}$ and $\bar{y}^e = \bar{y}$. In the experiment, the number of subjects per experimental economy is six. Evidence from other experiments indicates that four to six subjects are enough to justify the use of the competitive equilibrium as equilibrium concept (see e.g. Huck et al., 2004). Note, however, that also in a game theoretic analysis the unique Nash equilibrium is forecasting $\bar{\pi}$ and $\bar{y}$.

3.2 Course of Events and Implementation

The design is a between subject design with within session randomization. In the beginning, all participants are divided into groups (experimental economies) of six. Subjects only interact with other subjects in their group, without knowing who they are. The task subjects have is to make two-period-ahead forecasts of inflation and output gap. The average forecasts of all subjects in one group are then used to calculate the realizations of inflation and output gap according to model equations (1)–(3) (only the average forecasts $\bar{\pi}^e_{t+1}$ and $\bar{y}^e_{t+1}$ are needed to calculate the realizations π_t and y_t). The inflation target of the central bank in the experiment is $\bar{\pi} = 3.5$.[12] When

[11]The experiment can be seen as a controlled investigation of the outcomes of different monetary policies but also as a test between the rational and the behavioral models. While some people may argue that the best test of the models is to compare subjects' forecasts to the model predictions (in which the behavioral model does much better), others could question such a comparison on the ground that it is a within-treatment comparison; the directionally different hypotheses in our experiment make it a cleaner test (in laboratory experiments, the comparative statics of treatment comparisons are generally considered to be most robust and relevant; see Schram, 2005, Falk and Heckman, 2009, and Charness and Kuhn, 2011).

[12]This number has been chosen so as to be (i) large enough to have some distance from the zero lower bound as this is not supposed to be a liquidity trap experiment, (ii) different from focal points such as 2% or 2.5%, which are standard inflation targets in the real world, so that we can observe some

making their forecasts for period $t+1$, the information subjects can see on their screen (as numbers and partly also in graphs) is the following: all realizations of inflation, output gap, and interest rate up to period $t-1$, their own forecasts of inflation and output gap up to period t and their scores stating how close their past forecasts were to realized values up to period $t-1$ (these scores determine the payments). Figure 5 shows a screenshot of the experiment (a larger version of the same screenshot can be found in Appendix C).

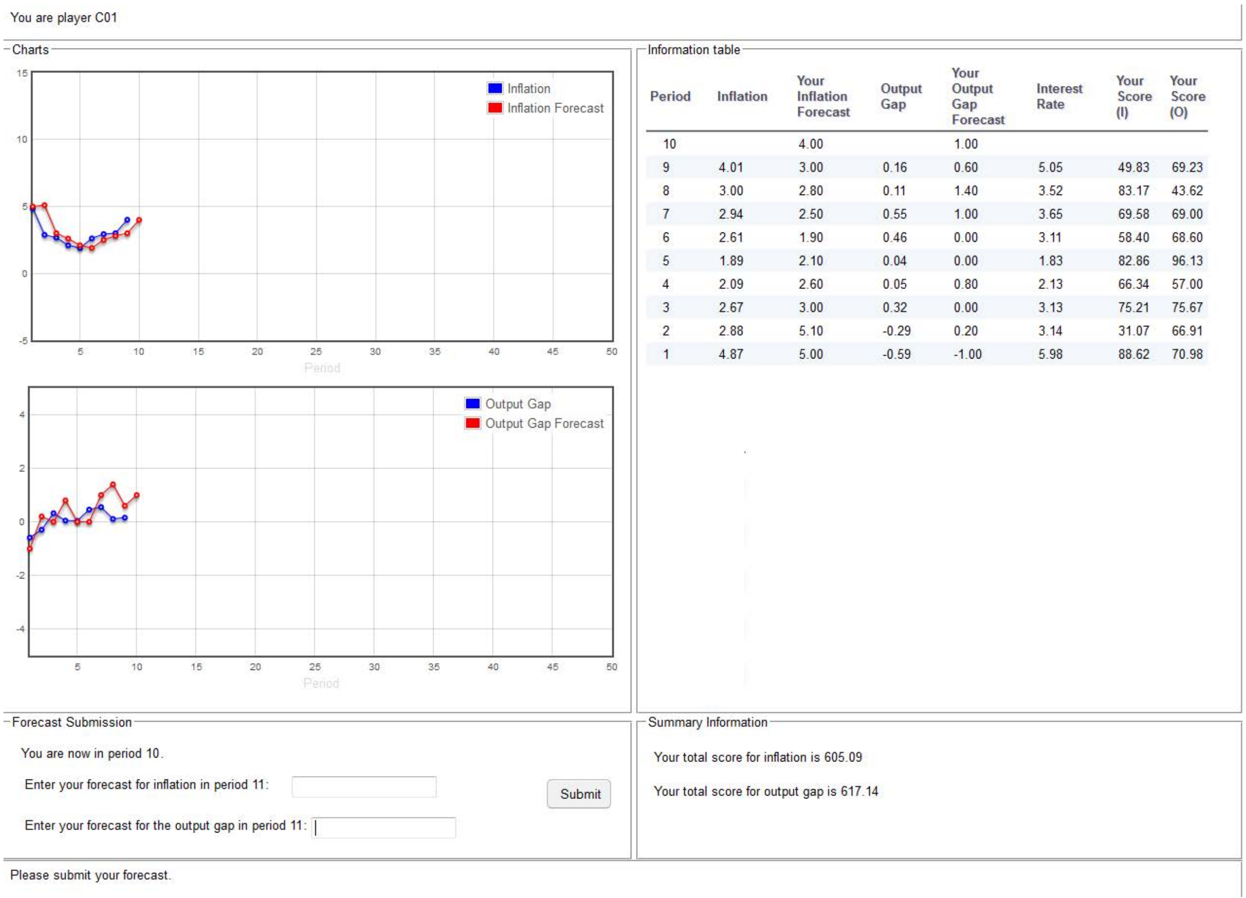


Figure 5: Screenshot

Subjects' payments depend on their forecasting performance. At the end of the experiment it is determined randomly for each participant whether she is paid for inflation forecasting or output gap forecasting. The total scores for inflation and output gap forecasting are the sums of the respective forecasting scores over all periods. This score is for subject i's inflation forecast in period t equal to $100/(1+|\pi^e_{t,i}-\pi_t|)$, where $\pi^e_{t,i}$ denotes subject i's forecast for period t and π_t the realized value of this period. The score for output gap forecasting is calculated analogously. This means that subjects'

learning in the experiment, and (iii) low enough so as not to be too far away from zero not to make the approximation from the log-linearized equations too imprecise.

payments decrease with the distance of the realizations from their forecasts.

In the instructions, subjects receive a qualitative description of the economy, describing the mechanisms governing the model equations. Concerning monetary policy, subjects in both treatments are only told that the central bank decreases the interest rate if it wants to increase inflation or output gap and that it increases the interest rate if it wants to decrease inflation or output gap.[13] Except for the precise formulation of the equations of the macroeconomic model, the instructions contain full information about the experiment (i.e. on the number of subjects per group, payments, etc.). The complete instructions can be found in Appendix B.

The experiment was programmed in java and conducted at the CREED laboratory at the University of Amsterdam. The experiment was conducted with 192 subjects recruited from the CREED subject pool (32 groups of six subjects each, distributed over nine sessions). After each session, participants had to fill out a short questionnaire. Participants were primarily undergraduate students, the average age was slightly above 22 years. About half of the participants were female, about two thirds were majoring in economics or business, and about two thirds were Dutch. During the experiment, 'points' were used as currency. These points were exchanged for euros at the end of each session at an exchange rate of 0.75 euros per 100 points. The experiment lasted around two hours, and participants earned on average 30.45 euros. The series of error terms used in the model equations (g_t and u_t in equations 1 and 2) were different from group to group within each treatment, but the sets of noise series in both treatments were the same.[14]

[13]As the experiment uses two-period ahead forecasts, subjects are asked after having finished the instructions to enter forecasts for periods 1 and 2 simultaneously. Subjects therefore receive some indication of reasonable values by being told in the instructions that in economies similar to the one at hand inflation has historically been between -5% and 10% and the output gap between -5% and 5%.

[14] Before conducting the experiment, two pilot sessions were conducted (with a total of six groups). The pilot sessions differ from the actual experiment as follows: the error terms added to the model equations had a larger standard deviation, a different inflation target was used, and subjects in the pilot did not receive any information on the number of participants in each group. For two of the groups also a different combination of parameters for the Taylor rule was used.

We exclude two of the groups from the analysis (including these two groups, the experiment was conducted with 204 subjects). One of the groups was excluded, because of a very large typo (30 instead of 3.0; the corresponding participant notified us about this typo in the post-experiment questionnaire). The other group was excluded due to severe misunderstandings of one subject, who systematically stayed very far from the actual realizations (thereby also losing a lot of money). Our conclusions do not change if we include these groups in our analysis. The realizations and forecasts of inflation and output gap for these two groups are shown in Figure 20, Appendix C.

3.3 Results

There are data of 32 different groups, 17 in $T1$ and 15 in $T2$. The groups' actions do not influence one another in any way, thus the observations at the group level are statistically independent. To get a good overview of the data, consider Figures 6 and 7 (the data for each group including all individual forecasts can be found in Appendix C).

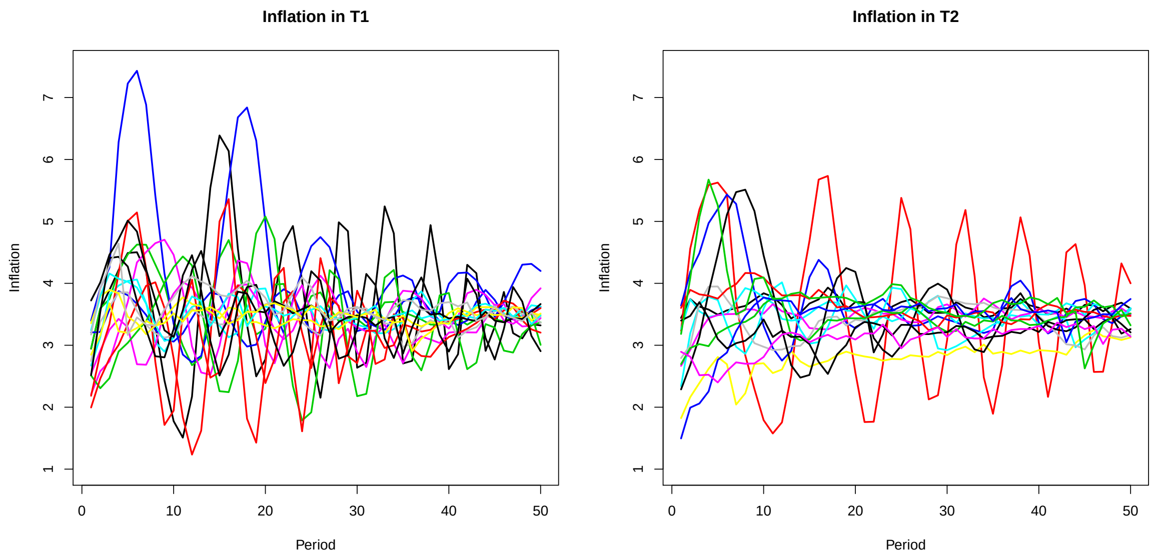


Figure 6: Realized inflation for all groups in both treatments

Notes: Each line represents realized inflation in one economy. On the horizontal axis is the number of periods (1 to 50), on the vertical axis inflation in percent (from 1 to 7.5).

Figure 6 gives an overview of inflation in all experimental economies, separately for $T1$ and $T2$. Each line corresponds to the inflation in one experimental economy, tracked over all 50 periods of the experiment. Almost all economies are close to the inflation target after 50 periods and for the economies with inflation still oscillating around the target the amplitude of these oscillations is decreasing. That many economies are converging to the steady state over the course of the experiment is not necessarily surprising as there are 50 periods without any changes to the underlying model (cf. Pfajfar and Zakelj, 2014, and Assenza et al., 2014b). The figure shows that inflation is indeed less volatile in $T2$ when also the output gap is targeted than in $T1$, as predicted by the behavioral model (consider for example inflation after half of the periods: Except for one economy, all economies in $T2$ already exhibit relatively low volatility and are close to the target, while in $T1$ many economies still exhibit wildly fluctuating inflation). Figure 7 shows the output gap for all experimental economies. Here, the differences are even larger, the output gap is much more volatile in $T1$ than in $T2$. This was to be expected, both models predict that the output gap is more stable

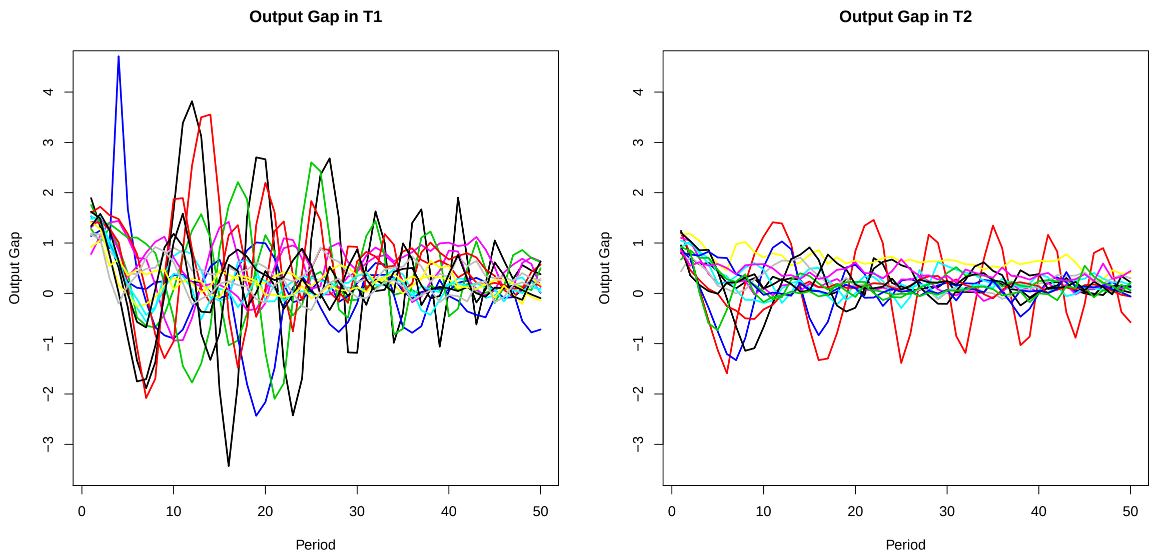


Figure 7: Realized output gap for all groups in both treatments

Notes: Each line represents realized output gap in one economy. On the horizontal axis is the number of periods (1 to 50), on the vertical axis output gap in percent (from −3.5 to 4.5). Each line has the same color as the line for the same group's inflation in Figure 6.

when it is also targeted by the central bank. While inflation and output gap volatility are quite different between the treatments, these variables generally fluctuate around their steady state values: The mean of inflation over all 50 periods is between 3.13 and 4.33 in $T1$ and between 2.79 and 3.76 in $T2$ for all groups, the mean of the output gap is between −0.12 and 0.70 in $T1$ and between 0.05 and 0.66 in $T2$.

We now turn to more detail about inflation volatility in the experiment. As in Section 2.3, we use $v(\pi) = \frac{1}{T-1}\sum_{t=2}^{T}(\pi_t - \pi_{t-1})^2$ as measure of inflation volatility (see Footnote 8). The values of this measure in all experimental economies can be seen in Figure 8 where the empirical cumulative distribution functions (ECDFs) are drawn, for groups in both treatments (for each value on the horizontal axis, the ECDF shows on the vertical axis the fraction of groups in each treatment with inflation volatility less or equal to this value; the colored dots stand for the actual observations). It can easily be seen that inflation volatility is lower in $T2$ than in $T1$. In fact, the whole ECDF of observations in $T2$ lies to the left of the ECDF of observations in $T1$ (the single one high value in $T2$, i.e. the rightmost blue dot, corresponds to the oscillating red line in the right graph of Figure 6).

In order to test the statistical significance of this finding we use a Wilcoxon rank-sum test. We test the null-hypothesis that inflation volatility is lower or equal in $T1$ than in $T2$ against the alternative hypothesis that inflation volatility is lower in $T2$.[15]

[15]Strictly speaking, the Wilcoxon rank-sum test tests the null-hypothesis that the distribution shifts

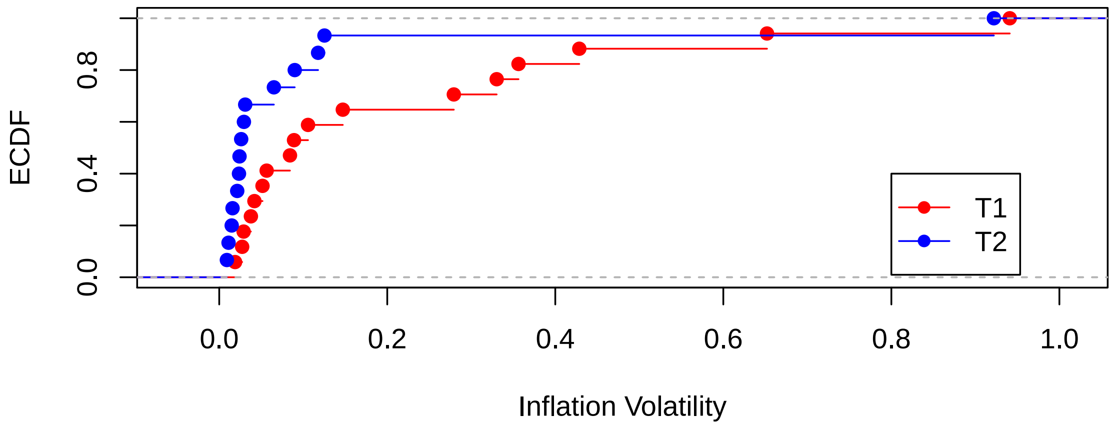


Figure 8: Empirical distribution functions of inflation volatility

Notes: For each value on the horizontal axis, the fraction of observations with inflation volatility less or equal to this value (i.e. the ECDF) is shown on the vertical axis, separately for $T1$ and $T2$.

This test rejects the null-hypothesis ($p = 0.006$). The Wilcoxon rank-sum test has the advantage that it makes very unrestrictive assumptions on the underlying data. Note, however, that the results are robust to employing different tests.[16]

4 Concluding Remarks

We have conducted a learning to forecast experiment to test the predictions of a macroeconomic model with behavioral expectations. This behavioral model yields results that are in contrast to the results from the same macroeconomic model based on rational expectations. Namely, the behavioral model yields that inflation volatility can be reduced if the central bank reacts to the output gap on top of inflation. The predictions of the behavioral model are supported by the outcomes of our learning to forecast experiment in which only the monetary policy reaction function of the central bank was modified as a treatment variable.

These results are relevant for monetary policy analysis and important for central banks. They give support to a trade-off between inflation and output-gap that is different than usually assumed based on the standard models with rational expectations. The policy implications are particularly straightforward for central banks only

to the right (from $T1$ to $T2$) or that it does not change.

[16]The data are not normally distributed, but the logarithms of the data look rather close to a normal distribution (and are statistically not significantly different from it according to a Kolmogorov-Smirnov test). A t-test on the logarithms of the data also rejects the null-hypothesis ($p = 0.009$).

aiming at price stability, such as for example the ECB; these banks should react to the output gap even if they are ultimately only interested in price stability.

References

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

Anufriev, M., Assenza, T., Hommes, C. H., and Massaro, D. (2013). Interest rate rules and macroeconomic stability under heterogeneous expectations. *Macroeconomic Dynamics*, 17:1574–1604.

Anufriev, M. and Hommes, C. H. (2012). Evolutionary selection of individual expectations and aggregate outcomes. *American Economic Journal: Microeconomics*, 4:35–64.

Arifovic, J. and Sargent, T. (2003). Laboratory experiments with an expectational phillips curve. In *Evolution and Procedures in Central Banking*, pages 23–55. Cambridge University Press.

Arthur, W. B., Holland, J. H., LeBaron, B., Palmer, R. G., and Tayler, P. (1997). Asset pricing under endogenous expectations in an artificial stock market. In Arthur, W. B., Durlauf, S., and Lane, D., editors, *The Economy as an Evolving Complex System II*. Addison-Wesley.

Assenza, T., Bao, T., Hommes, C., and Massaro, D. (2014a). Experiments on expectations in macroeconomics and finance. In Duffy, J., editor, *Experiments in Macroeconomics*, volume 17 of *Research in Experimental Economics*.

Assenza, T., Heemeijer, P., Hommes, C., and Massaro, D. (2014b). Managing self-organization of expectations through monetary policy: a macro experiment. CeNDEF Working Paper 14-07, University of Amsterdam.

Bao, M., Hommes, C., Sonnemans, J., and Tuinstra, J. (2012). Individual expectations, limited rationality and aggregate outcomes. *Journal of Economic Dynamics and Control*, 36:1101–1120.

Benhabib, J., Evans, G. W., and Honkapohja, S. (2014). Liquidity traps and expectation dynamics: Fiscal stimulus or fiscal austerity? *Journal of Economic Dynamics and Control*, 45:220–238.

Bernasconi, M. and Kirchkamp, O. (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.

Bloomfield, R. and Hales, J. (2002). Predicting the next step of a random walk: Experimental evidence of regime-switching beliefs. *Journal of Financial Economics*, 65:397–414.

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

Branch, W. and McGough, B. (2009). A new keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control*, 33:1036–1051.

Branch, W. and McGough, B. (2010). Dynamic predictors selection in a new keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control*, 34(8):1492–1508.

Branch, W. A., Carlson, J., Evans, G. W., and McGough, B. (2009). Monetary policy, endogenous inattention and the volatility trade-off. *The Economic Journal*, 119(534):123–157.

Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica*, 65(5):1059–1095.

Brock, W. A. and Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22:1235–1274.

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

Camerer, C. F., Loewenstein, G., and Rabin, M. (2011). *Advances in behavioral economics*. Princeton University Press.

Carroll, C. (2003). Macroeconomic expectations of households and professional forecasters. *Quarterly Journal of economics*, 118:269–298.

Charness, G. and Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? *Handbook of Labor Economics*, 4:229–330.

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

Cornand, C. and M'Baye, C. K. (2013). Does Inflation Targeting Matter? An Experimental Investigation. Technical Report 30, University of Lyon – GATE.

Davis, D. and Korenok, O. (2011). Nominal shocks in monopolistically competitive markets: An experiment. *Journal of Monetary Economics*, 58(6):578–589.

De Grauwe, P. (2011). Animal spirits and monetary policy. *Economic theory*, 47(2-3):423–457.

De Grauwe, P. (2012a). Booms and busts in economic activity: A behavioral explanation. *Journal of Economic Behavior & Organization*, 83(3):484–501.

De Grauwe, P. (2012b). *Lectures on behavioral macroeconomics*. Princeton University Press, New Jersey.

De Grauwe, P. and Kaltwasser, P. R. (2012). Animal spirits in the foreign exchange market. *Journal of Economic Dynamics and Control*, 36(8):1176–1192.

Diks, C. and van der Weide, R. (2005). Herding, a-synchronous updating and heterogeneity in memory in a cbs. *Journal of Economic Dynamics and Control*, 29:741–763.

Duffy, J. (2012). Macroeconomics: A survey of laboratory research. Working Paper 334, University of Pittsburgh.

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

Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952):535–538.

Frankel, J. and Froot, K. (1987). Using survey data to test standard propositions regarding exchange rate expectations. *American Economic Review*, 77(1):133–153.

Frankel, J. and Froot, K. (1991). Chartists, fundamentalists and the demand for dollars. In Courakis, A. and Taylor, M., editors, *Private Behaviour and Government Policy in Interdependent Economies*. Oxford: Clarendon.

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

Gigerenzer, G. and Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. Mit Press.

Gigerenzer, G. and Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press, USA.

Grether, D. and Plott, C. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69(4):623–638.

Guesnerie, R. (2009). Macroeconomic and Monetary Policies from the Eductive Viewpoint. In *Monetary Policy under Uncertainty and Learning*, volume 13 of *Central Banking, Analysis, and Economic Policies Book Series*, chapter 6, pages 171–202. Central Bank of Chile.

Heemeijer, P., Hommes, C. H., Sonnemans, J., and Tuinstra, J. (2009). Price stability and volatility in markets with positive and negative expectations feedbacks: An experimental investigation. *Journal of Economic Dynamics and Control*, 33:1052–1072.

Hommes, C. H. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control*, 35:1–24.

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

Hommes, C. H., Sonnemans, J., Tuinstra, J., and van de Velden, H. (2005b). Coordination of expectations in asset pricing experiments. *Review of Financial Studies*, 18(3):955–980.

Hommes, C. H., Sonnemans, J., Tuinstra, J., and van de Velden, H. (2008). Expectations and bubbles in asset pricing experiments. *Journal of Economic Behavior and Organization*, 67(1):116–133.

Huck, S., Normann, H., and Oechssler, J. (2004). Two are few and four are many: number effects in experimental oligopolies. *Journal of Economic Behavior and Organization*, 53:435–446.

Kahneman, D., Knetsch, J., and Tversky, A. (1991). Anomalies: The endowment effect, loss aversion and status quo bias. *Journal of Economic Perspectives*, 5(1):193–206.

Kelley, H. and Friedman, D. (2002). Learning to forecast price. *Economic Inquiry*, 40(4):556–573.

Kryvtsov, O. and Petersen, L. (2013). Expectations and monetary policy: Experimental evidence. Working Papers 13-44, Bank of Canada.

Kurz, M. (2011). A new keynesian model with diverse beliefs. Technical report, Stanford University.

Kurz, M., Piccillo, G., and Wu, H. (2013). Modeling diverse expectations in an aggregated new keynesian model. *Journal of Economic Dynamics and Control*, 37:1403–1433.

Lei, V. and Noussair, C. N. (2002). An experimental test of an optimal growth model. *American Economic Review*, pages 549–570.

Mankiw, N., Reis, R., and Wolfers, J. (2003). Disagreement about inflation expectations. In Gertler, M. and Rogoff, K., editors, *NBER Macroeconomics Annual*.

Manski, C. and McFadden, D. (1981). Alternative estimators and sample designs for discrete choice analysis. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data*. MIT Press.

Marcet, A. and Nicolini, J. P. (2003). Recurrent hyperinflations and learning. *The American Economic Review*, 93(5):1476–1498.

Marimon, R. and Sunder, S. (1993). Indeterminacy of equilibria in a hyperinflationary world: Experimental evidence. *Econometrica*, 61:1073–1107.

Massaro, D. (2013). Heterogeneous expectations in monetary dsge models. *Journal of Economic Dynamics and Control*, 37(3):680–692.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.

Pfajfar, D. and Zakelj, B. (2014). Experimental evidence on inflation expectations formation. *Journal of Economic Dynamics and Control*, 44:147–168.

Schram, A. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12(2):225–237.

Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185:1124–1130.

Tversky, A. and Thaler, R. (1990). Anomalies: Preference reversal. *Journal of Economic Perspectives*, 4(2):201–211.

Van Huyck, J. B., Cook, J. P., and Battalio, R. C. (1994). Selection dynamics, asymptotic stability, and adaptive behavior. *Journal of Political Economy*, pages 975–1005.

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

Woodford, M. (2010). Robustly optimal monetary policy with near-rational expectations. *American Economic Review*, 100(1):274–303.

Woodford, M. (2013). Macroeconomic Analysis Without the Rational Expectations Hypothesis. *Annual Review of Economics*, 5(1):303–346.

A Appendix (for Online Publication): Additional Graphs from Simulations of the Macroeconomic Model

Figure 9 shows inflation volatility as a function of the output gap reaction coefficient ϕ_y for the model assuming rational expectations, similarly to Figure 1a. The graph now shows multiple coefficients of ϕ_π simultaneously (from top to bottom the lines correspond to ϕ_π-values of 1.4, 1.5, 1.6, and 1.7). Figure 10 shows the same graph for the behavioral model (again the lines correspond to ϕ_π-values of 1.4, 1.5, 1.6, and 1.7, from top to bottom).

Figures 11 and 12 show graphs similar to Figure 1b for different combinations of starting values of inflation and output gap (i.e. inflation and output gap are set to these starting values in the first two periods). In all cases the U-shape arises similarly to Figure 1b.

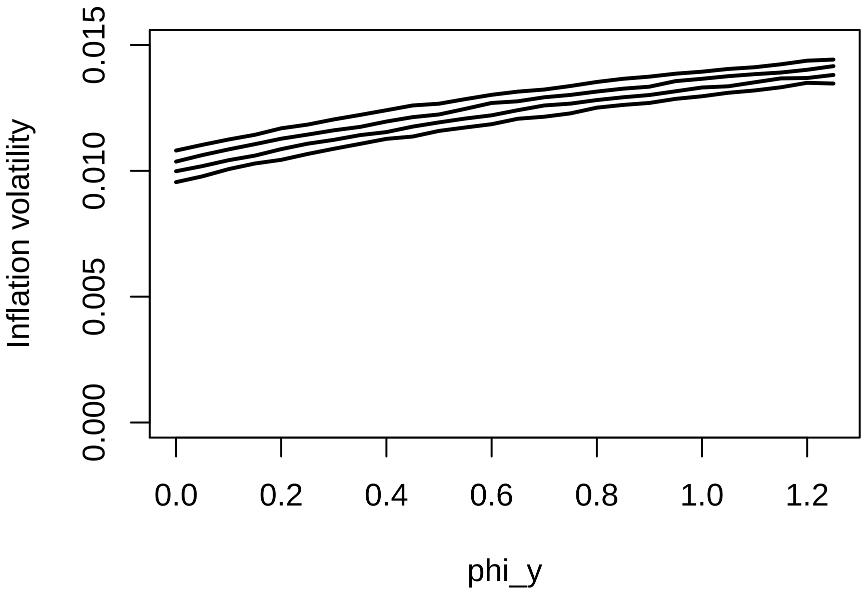


Figure 9: Inflation volatility as a function of ϕ_y for the rational model for different values of ϕ_π (from 1.4 (top line) to 1.7)

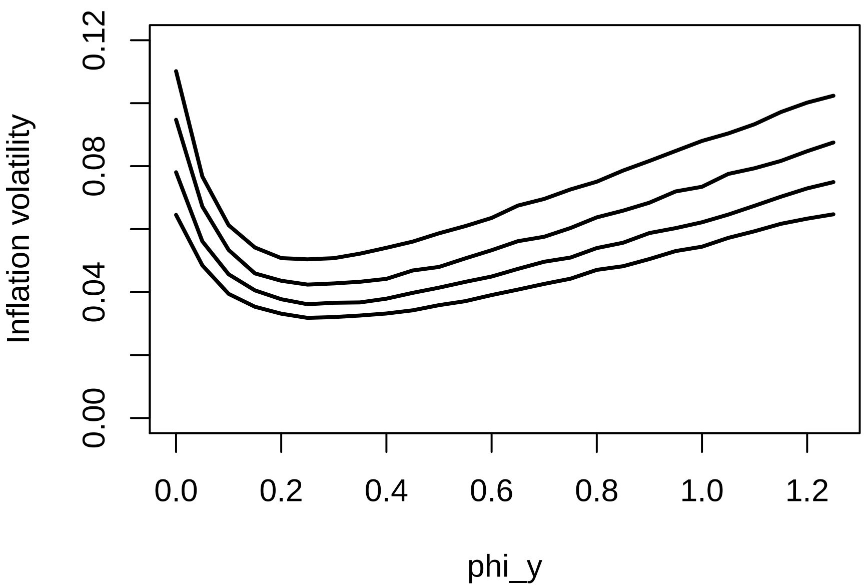
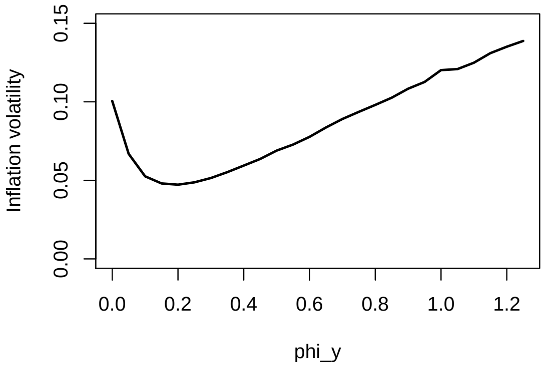
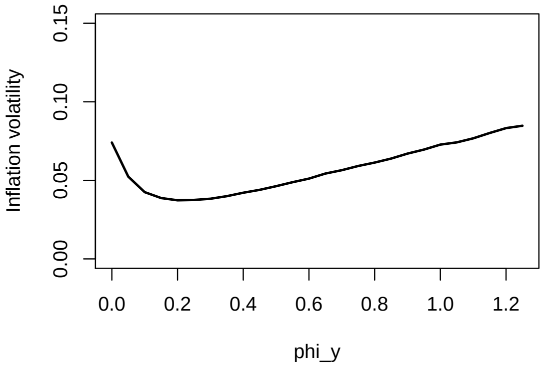


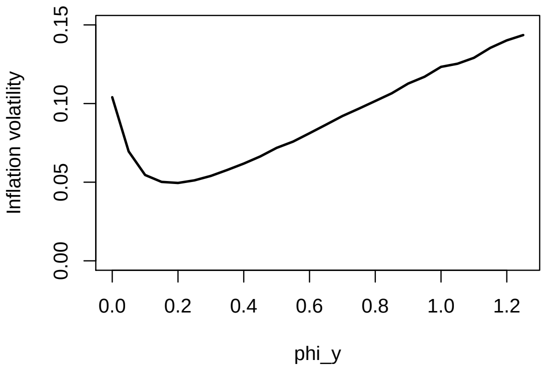
Figure 10: Inflation volatility as a function of ϕ_y for the behavioral model for different values of ϕ_π (from 1.4 (top line) to 1.7)



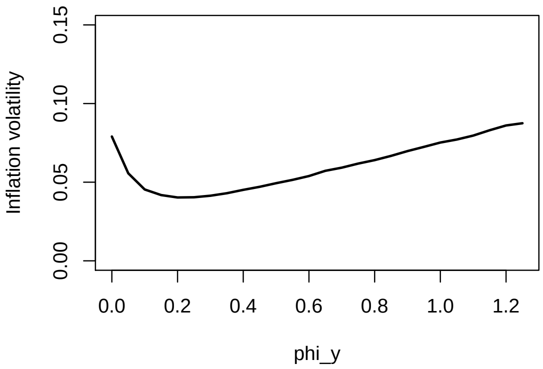
(a) $\pi_{start} = 2.5$, $y_{start} = -0.5$



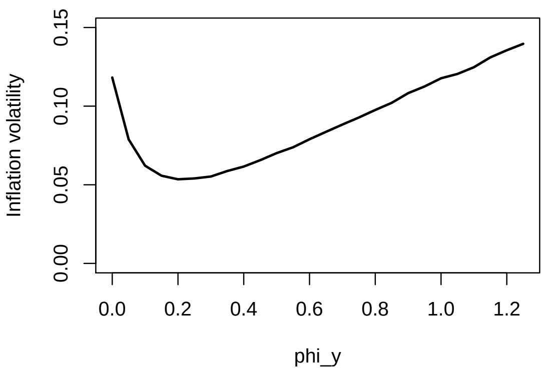
(b) $\pi_{start} = 3.0$, $y_{start} = -0.5$



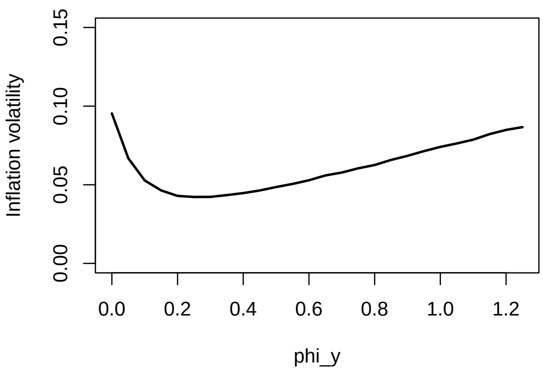
(c) $\pi_{start} = 2.5$, $y_{start} = 0$



(d) $\pi_{start} = 3.0$, $y_{start} = 0$



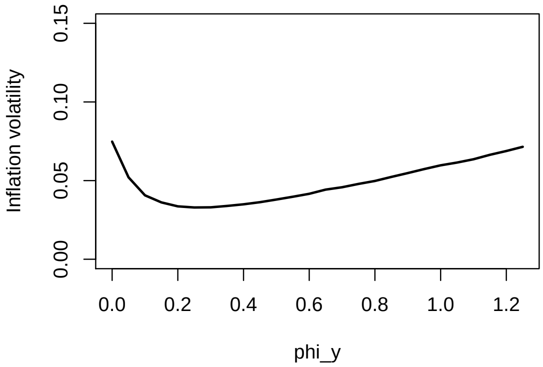
(e) $\pi_{start} = 2.5$, $y_{start} = 0.5$



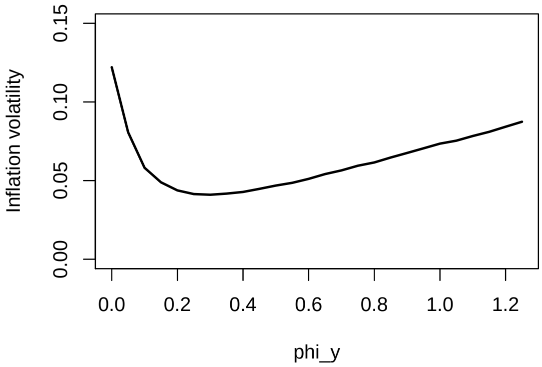
(f) $\pi_{start} = 3.0$, $y_{start} = 0.5$

Figure 11: Inflation volatility in the behavioral model for different starting values

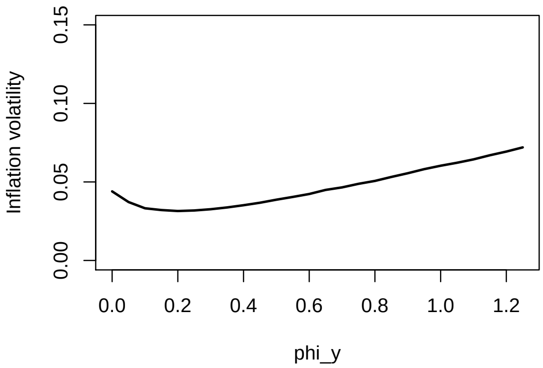
Notes: This figure shows the effect of parameter ϕ_y on inflation volatility for different starting values of y and π ($\phi_\pi = 1.5$ throughout).



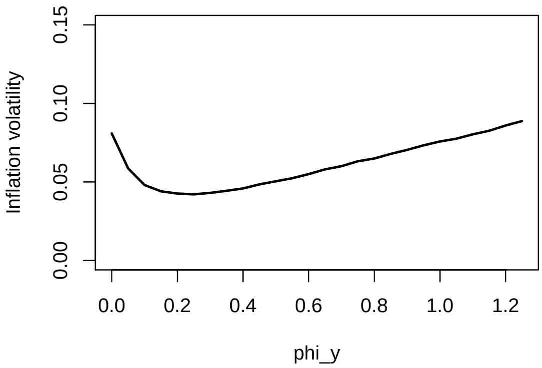
(a) $\pi_{start} = 3.5$, $y_{start} = -0.5$



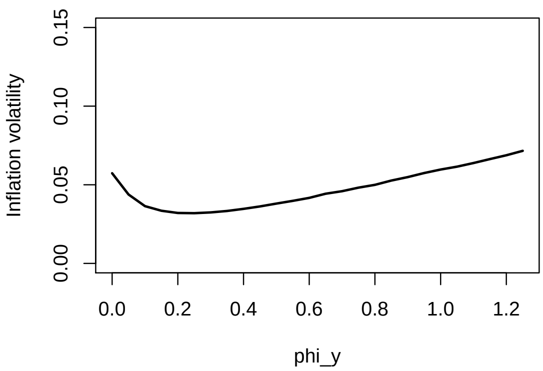
(b) $\pi_{start} = 4.0$, $y_{start} = -0.5$



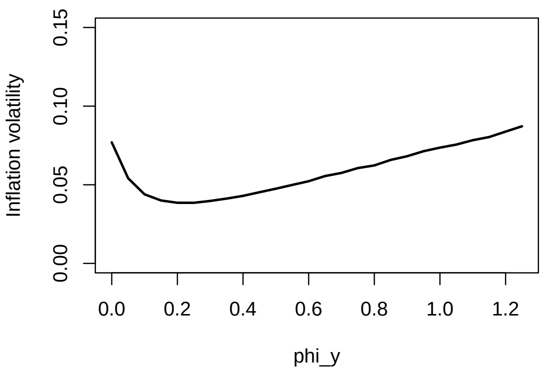
(c) $\pi_{start} = 3.5$, $y_{start} = 0$



(d) $\pi_{start} = 4.0$, $y_{start} = 0$



(e) $\pi_{start} = 3.5$, $y_{start} = 0.5$



(f) $\pi_{start} = 4.0$, $y_{start} = 0.5$

Figure 12: Inflation volatility in the behavioral model for different starting values

Notes: This figure shows the effect of parameter ϕ_y on inflation volatility for different starting values of y and π ($\phi_\pi = 1.5$ throughout).

B Appendix (for Online Publication): Instructions in the Experiment

Subjects in the experiment received the following instructions (as subjects only received qualitative information on the model governing the experimental economy the instructions are the same for both treatments):

Instructions

Welcome to this experiment! The experiment is anonymous, the data from your choices will only be linked to your station ID, not to your name. You will be paid privately at the end, after all participants have finished the experiment. After the main part of the experiment and before the payment you will be asked to fill out a short questionnaire. On your desk you will find a calculator and scratch paper, which you can use during the experiment.

During the experiment you are not allowed to use your mobile phone. You are also not allowed to communicate with other participants. If you have a question at any time, please raise your hand and someone will come to your desk.

General information and experimental economy

All participants will be randomly divided into groups of six people. The group composition will not change during the experiment. You and all other participants will take the roles of statistical research bureaus making predictions of inflation and the so-called "output gap". The experiment consists of 50 periods in total. In each period you will be asked to predict inflation and output gap for the next period. The economy you are participating in is described by three variables: inflation π_t, output gap y_t and interest rate i_t. The subscript t indicates the period the experiment is in. In total there are 50 periods, so t increases during the experiment from 1 to 50.

Inflation

Inflation measures the percentage change in the price level of the economy. In each period, inflation depends on inflation predictions of the statistical research bureaus in the economy (a group of six participants in this experiment), on actual output gap and on a random term. There is a positive relation between the actual inflation and both inflation predictions and actual output gap. This means for example that if the inflation predictions of the research bureaus increase, then actual inflation will also increase (everything else equal). In economies similar to this one, inflation has historically been between −5% and 10%.

Output gap

The output gap measures the percentage difference between the Gross Domestic Product (GDP) and the natural GDP. The GDP is the value of all goods produced during a period in the economy. The natural GDP is the value the total production would have if prices in the economy were fully flexible. If the output gap is positive (negative), the economy therefore produces more (less) than the natural GDP. In each period the output gap depends on inflation predictions and output gap predictions of the statistical bureaus, on the interest rate and on a random term. There is a positive relation between the output gap and inflation predictions and also between the output gap and output gap predictions. There is a negative relation between the output gap and the interest rate. In economies similar to this one, the output gap has historically been between −5% and 5%.

Interest Rate

The interest rate measures the price of borrowing money and is determined by the central bank. If the central bank wants to increase inflation or output gap it decreases the interest rate, if it wants to decrease inflation or output gap it increases the interest rate.

Prediction task

Your task in each period of the experiment is to predict inflation and output gap in the next period. When the experiment starts, you have to predict inflation and output gap for the first two periods, i.e. π_1^e and π_2^e, and y_1^e and y_2^e. The superscript e indicates that these are predictions. When all participants have made their predictions for the first two periods, the actual inflation (π_1), the actual output gap (y_1) and the interest rate (i_1) for period 1 are announced. Then period 2 of the experiment begins. In period 2 you make inflation and output gap predictions for period 3 (π_3^e and y_3^e). When all participants have made their predictions for period 3, inflation (π_2), output gap (y_2), and interest rate (i_2) for period 2 are announced. This process repeats itself for 50 periods.

Thus, in a certain period t when you make predictions of inflation and output gap in period $t+1$, the following information is available to you:

- Values of actual inflation, output gap and interest rate up to period $t-1$;
- Your predictions up to period t;

- Your prediction scores up to period $t-1$.

Payments

Your payment will depend on the accuracy of your predictions. You will be paid either for predicting inflation or for predicting the output gap. The accuracy of your predictions is measured by the absolute distance between your prediction and the actual values (this distance is the prediction error). For each period the prediction error is calculated as soon as the actual values are known; you subsequently get a prediction score that decreases as the prediction error increases. The table below gives the relation between the prediction error and the prediction score. The prediction error is calculated in the same way for inflation and output gap.

Prediction error	0	1	2	3	4	9
Score	100	50	33.33	25	20	10

Example: If (for a certain period) you predict an inflation of 2%, and the actual inflation turns out to be 3%, then you make an absolute error of $3\% - 2\% = 1\%$. Therefore you get a prediction score of 50. If you predict an inflation of 1%, and the actual inflation turns out to be negative 2% (i.e. -2%), you make a prediction error of $1\% - (-2\%) = 3\%$. Then you get a prediction score of 25. For a perfect prediction, with a prediction error of zero, you get a prediction score of 100. The figure below shows the relation between your prediction score (vertical axis) and your prediction error (horizontal axis). Points in the graph correspond to the prediction scores in the previous table.

[Figure 13 appears here in the experimental instructions.]

At the end of the experiment, you will have two total scores, one for inflation predictions and one for output gap predictions. These total scores simply consist of the sum of all prediction scores you got during the experiment, separately for inflation and output gap predictions. **When the experiment has ended, one of the two total scores will be randomly selected for payment.**

Your final payment will consist of 0.75 **euro for each** 100 **points in the selected total score (**200 **points therefore equals** 1.50 **euro). This will be the only payment from this experiment, i.e. you will not receive a show-up fee on top of it.**

Computer interface

The computer interface will be mainly self-explanatory. The top right part of the screen will show you all of the information available up to the period that you are in (in period t, i.e. when you are asked to make your prediction for period $t+1$, this will be actual inflation, output gap, and interest rate until period $t-1$, your predictions until period t, and the prediction scores arising from your predictions until period $t-1$ for both inflation (I) and output gap (O)). The top left part of the screen will show you the information on inflation and output gap in graphs. The axis of a graph shows values in percentage points (i.e. 3 corresponds to 3%). **Note that the values on the vertical axes may change during the experiment and that they are different between the two graphs – the values will be such that it is comfortable for you to read the graphs.**

In the bottom left part of the screen you will be asked to enter your predictions. When submitting your prediction, use a decimal point if necessary (not a comma). For example, if you want to submit a prediction of 2.5% **type** "2.5"**; for a prediction of** -1.75% **type** "-1.75"**.** The sum of the prediction scores over the different periods are shown in the bottom right of the screen, separately for your inflation and output gap predictions.

At the bottom of the screen there is a status bar telling you when you can enter your predictions and when you have to wait for other participants.

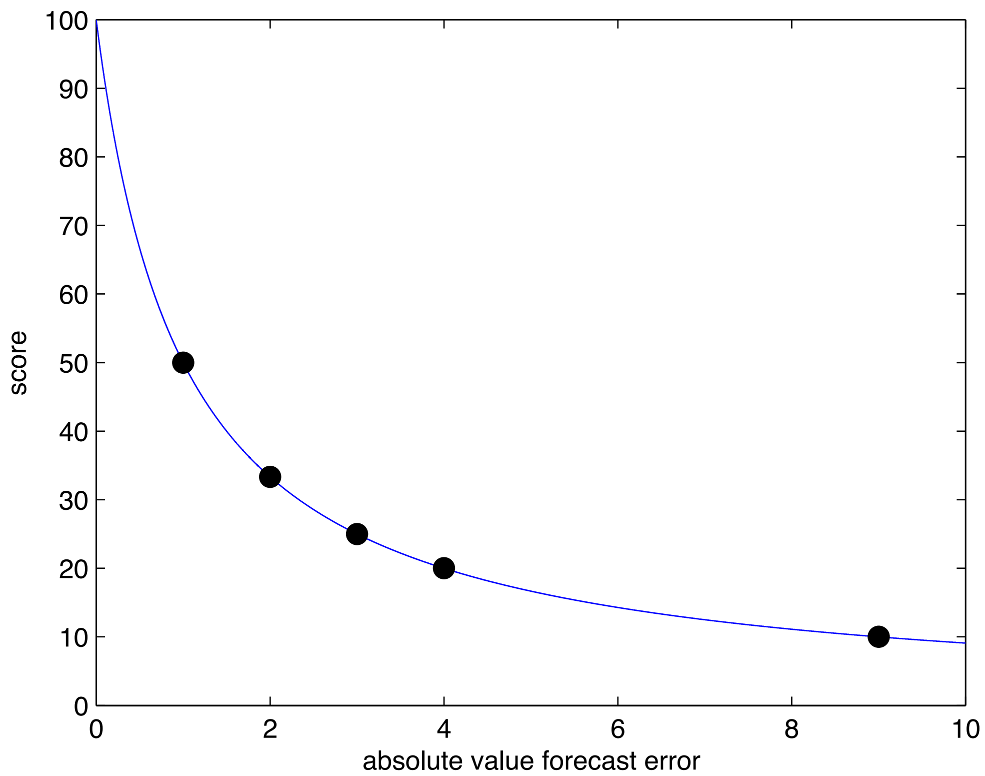


Figure 13: Relation score and forecast error (not labeled in the instructions)

C Appendix (for Online Publication): Additional Graphs of the Experimental Data and Screenshot

Figures 14 to 20 show the realizations and forecasts of inflation and output gap. Each graph corresponds to one group of six people (one experimental economy). The thick black line shows the realization of inflation, the thin dashed black lines show the inflation forecasts of the six individuals in the group. The thick gray line shows the realization of the output gap and the thin dashed gray lines show the output gap forecasts of all individuals in a group. On the horizontal axis are the periods (from 1 to 50), on the vertical axis are the values of inflation and output gap in percent (the numbers on the vertical axis reach from -3 to 8). The upper red line corresponds to the steady state value of inflation ($\bar{\pi} = 3.5$), the lower red line corresponds to the steady state value of the output gap ($\bar{y} = 0.1166667$). Figures 14 to 16 show all groups of treatment $T1$, Figures 17 to 19 show the groups of treatment $T2$. Figure 20 shows the two groups (from $T2$) that have been excluded from the analysis as explained in Footnote 14.

Figure 21 shows a screenshot (a larger version of the screenshot already used in Figure 5.

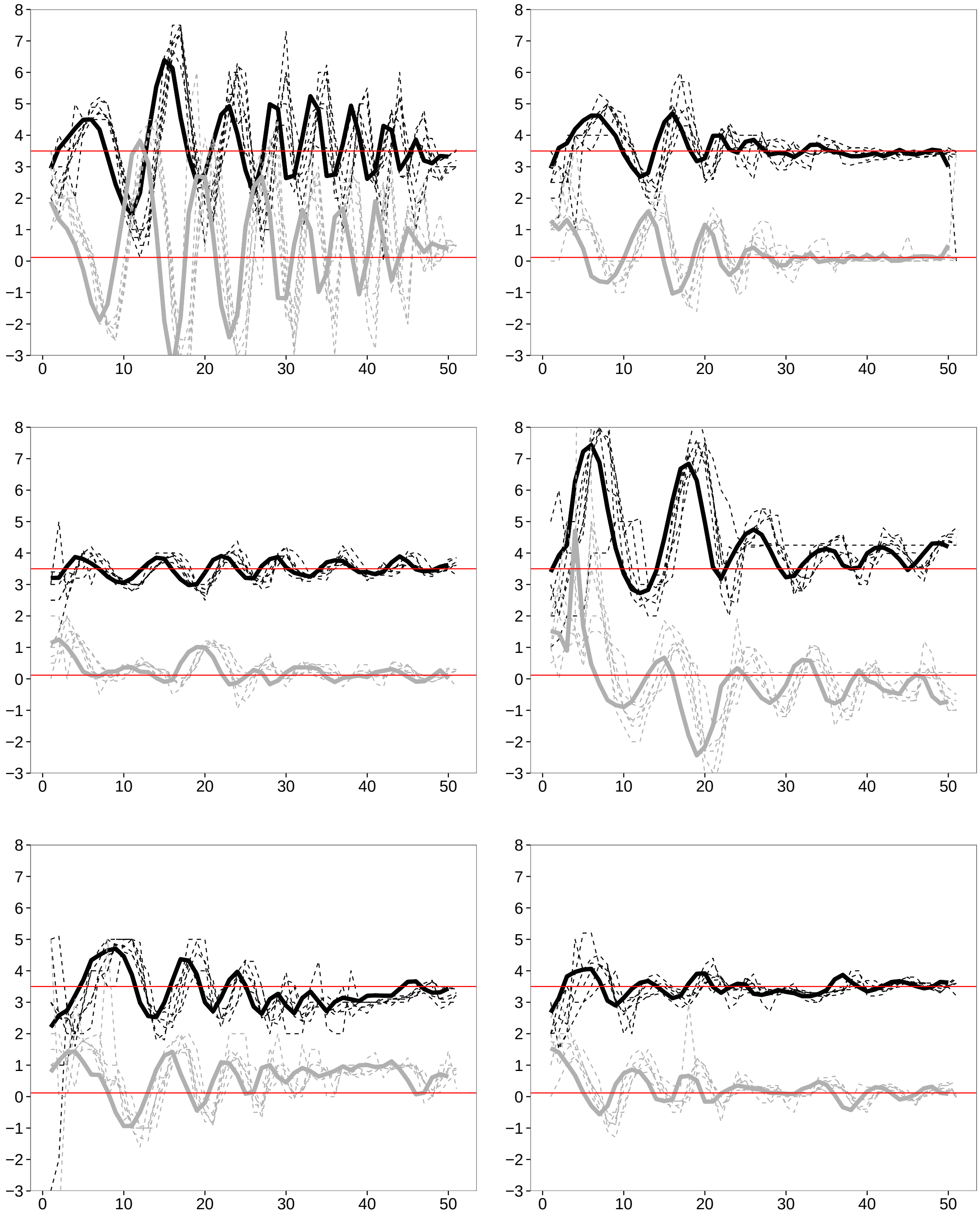


Figure 14: Realizations and forecasts of inflation and output gap ($T1$, groups $1-6$)

Notes: Each of the graphs corresponds to one group and shows realized inflation (thick black line), individual inflation forecasts (dashed black lines), realized output gap (thick gray line), and individual output gap forecasts (dashed gray lines) over the 50 periods of the experiment.

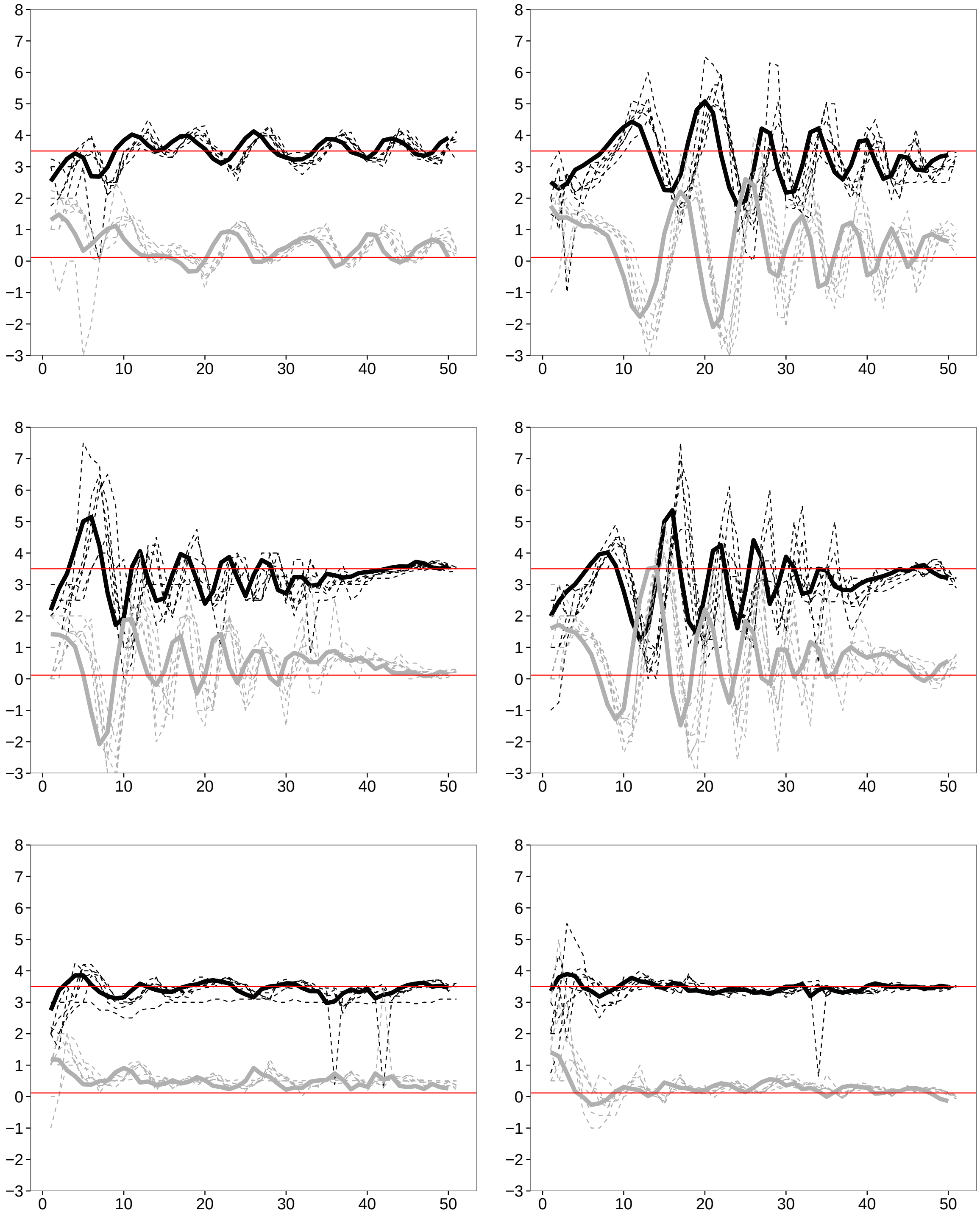


Figure 15: Realizations and forecasts of inflation and output gap ($T1$, groups $7-12$)

Notes: Each of the graphs corresponds to one group and shows realized inflation (thick black line), individual inflation forecasts (dashed black lines), realized output gap (thick gray line), and individual output gap forecasts (dashed gray lines) over the 50 periods of the experiment.

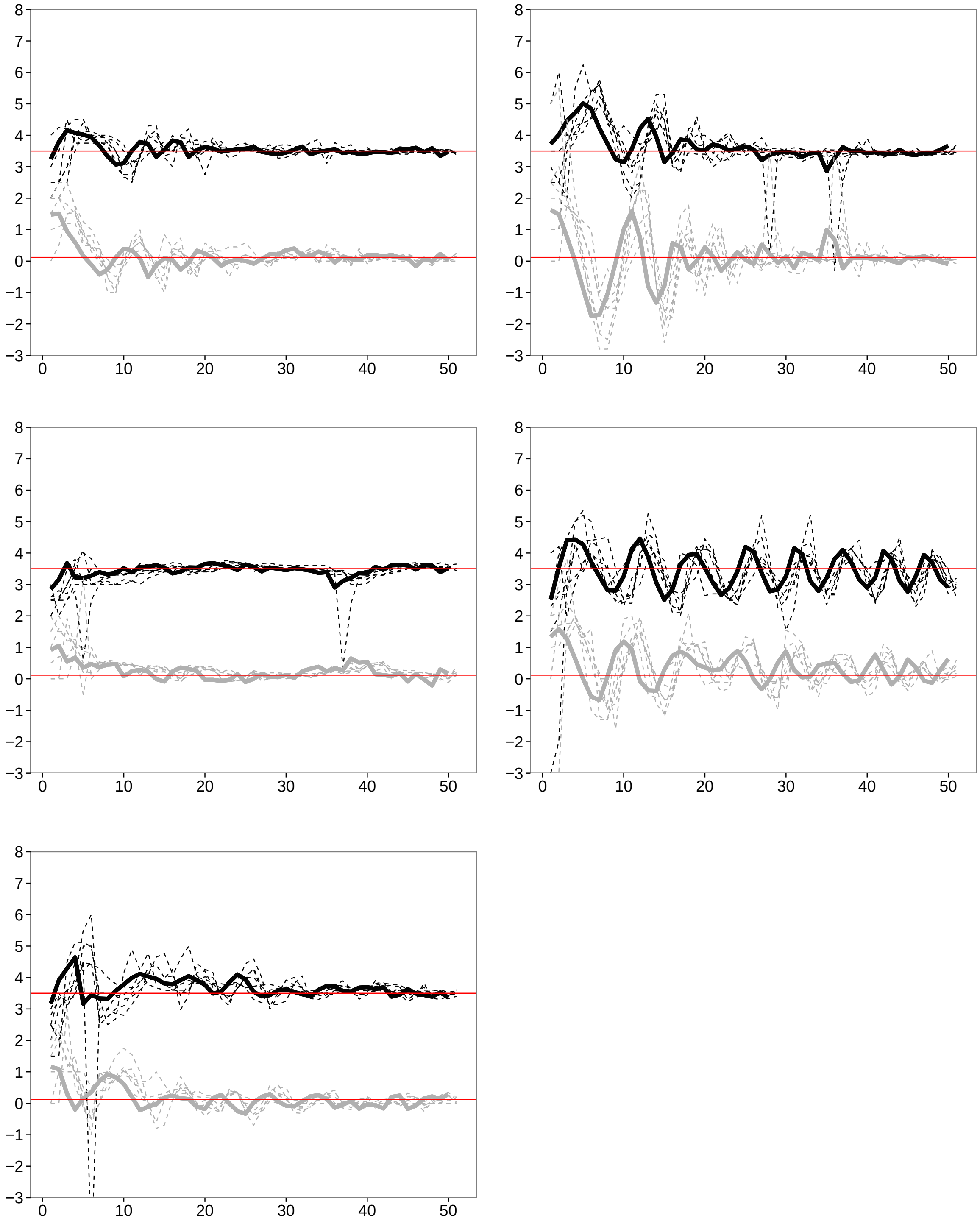


Figure 16: Realizations and forecasts of inflation and output gap ($T1$, groups $13-17$)

Notes: Each of the graphs corresponds to one group and shows realized inflation (thick black line), individual inflation forecasts (dashed black lines), realized output gap (thick gray line), and individual output gap forecasts (dashed gray lines) over the 50 periods of the experiment.

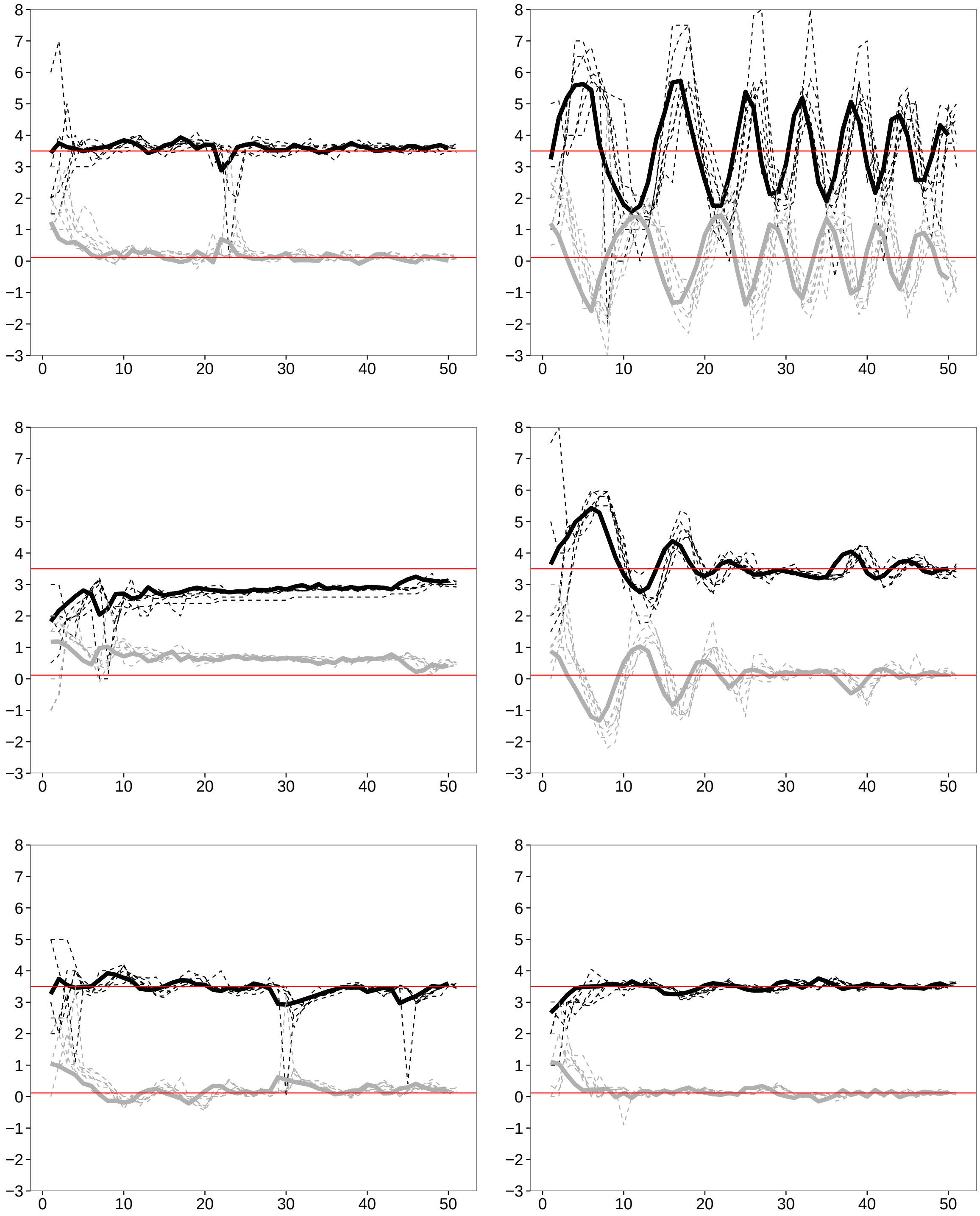


Figure 17: Realizations and forecasts of inflation and output gap ($T2$, groups $1-6$)

Notes: Each of the graphs corresponds to one group and shows realized inflation (thick black line), individual inflation forecasts (dashed black lines), realized output gap (thick gray line), and individual output gap forecasts (dashed gray lines) over the 50 periods of the experiment.

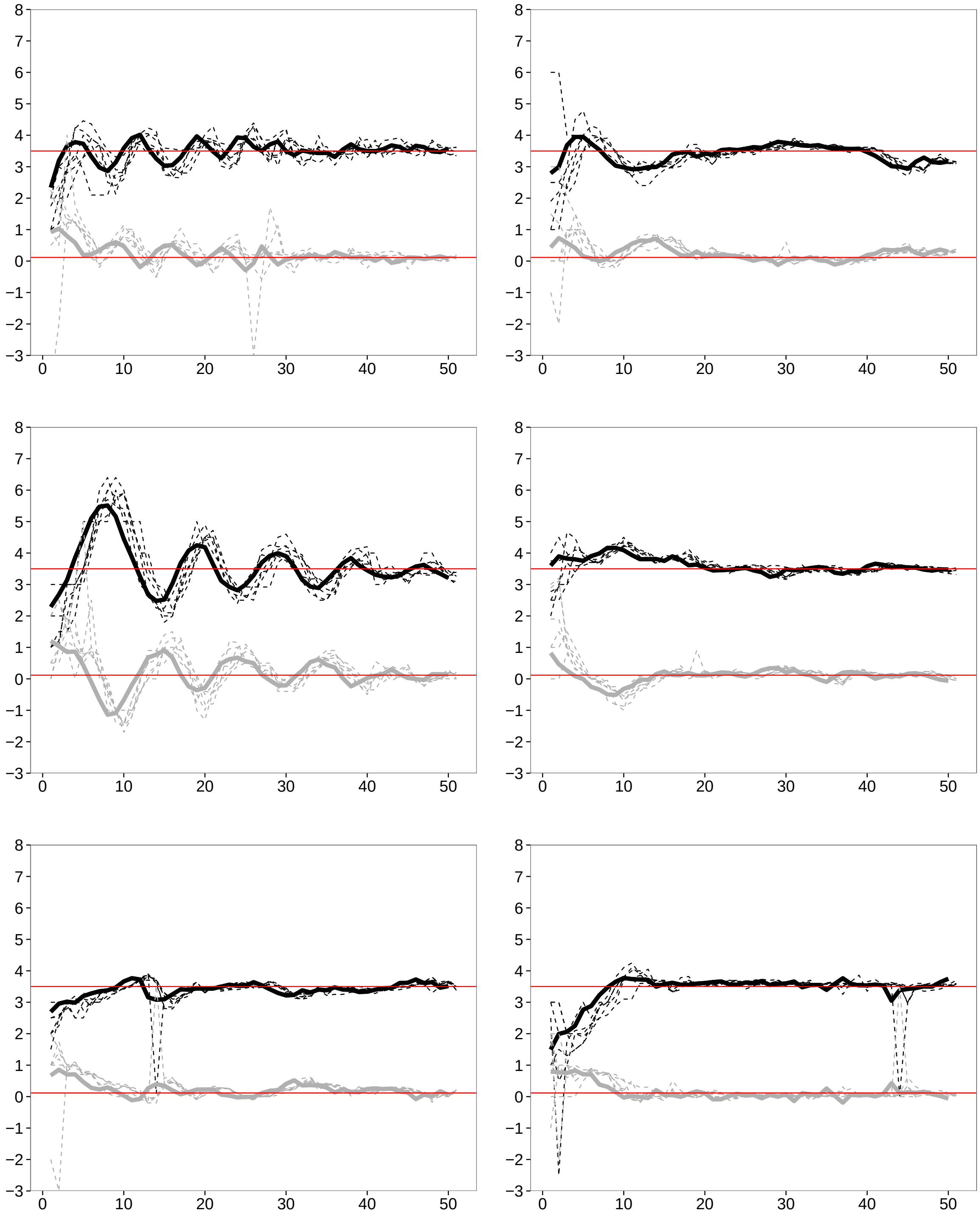


Figure 18: Realizations and forecasts of inflation and output gap ($T2$, groups $7-12$)

Notes: Each of the graphs corresponds to one group and shows realized inflation (thick black line), individual inflation forecasts (dashed black lines), realized output gap (thick gray line), and individual output gap forecasts (dashed gray lines) over the 50 periods of the experiment.

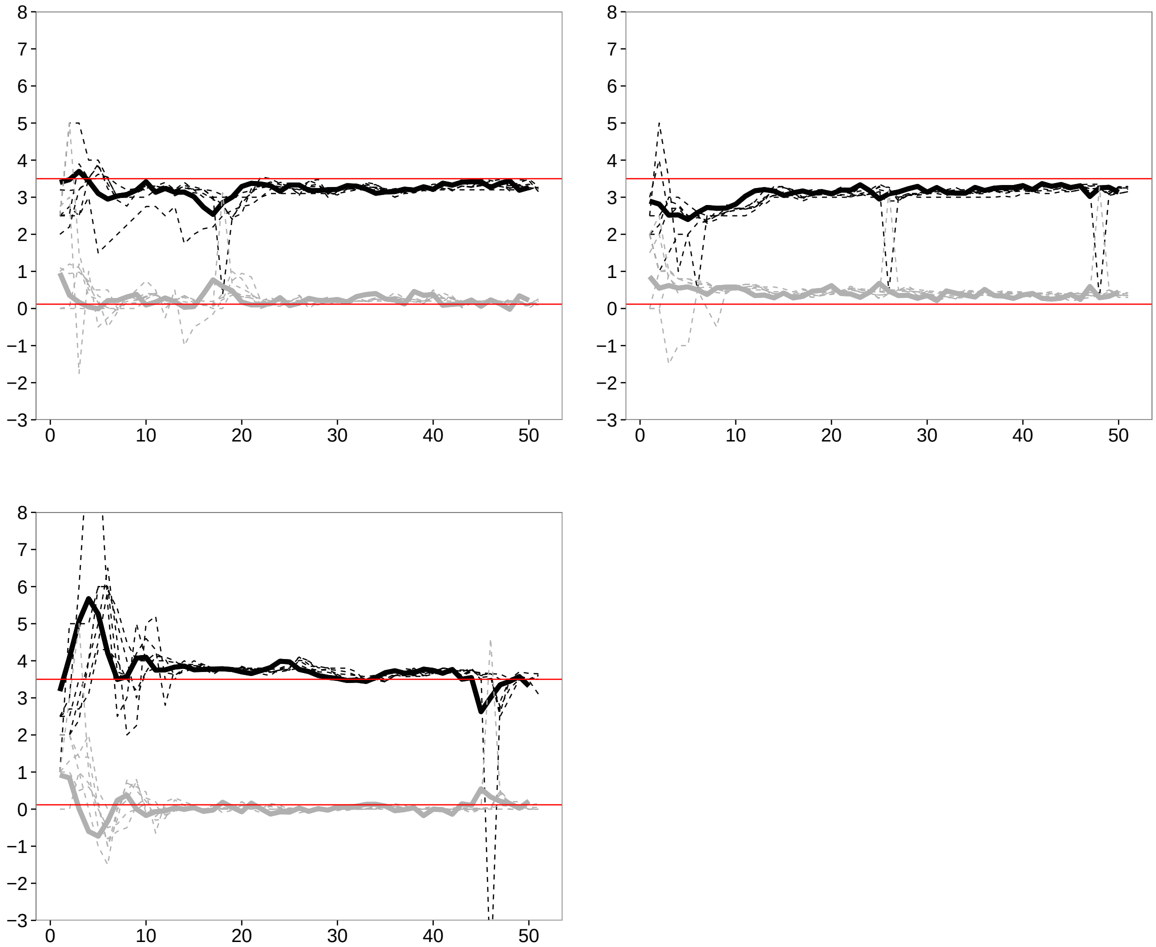


Figure 19: Realizations and forecasts of inflation and output gap ($T2$, groups $13-15$)

Notes: Each of the graphs corresponds to one group and shows realized inflation (thick black line), individual inflation forecasts (dashed black lines), realized output gap (thick gray line), and individual output gap forecasts (dashed gray lines) over the 50 periods of the experiment.

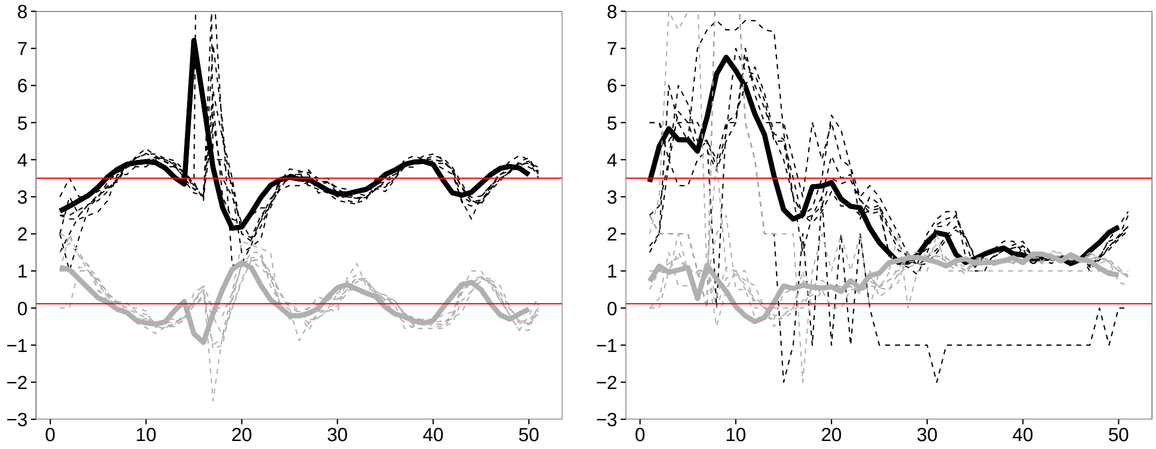


Figure 20: Realizations and forecasts of inflation and output gap (excluded groups)

Notes: Each of the graphs corresponds to one group and shows realized inflation (thick black line), individual inflation forecasts (dashed black lines), realized output gap (thick gray line), and individual output gap forecasts (dashed gray lines) over the 50 periods of the experiment.

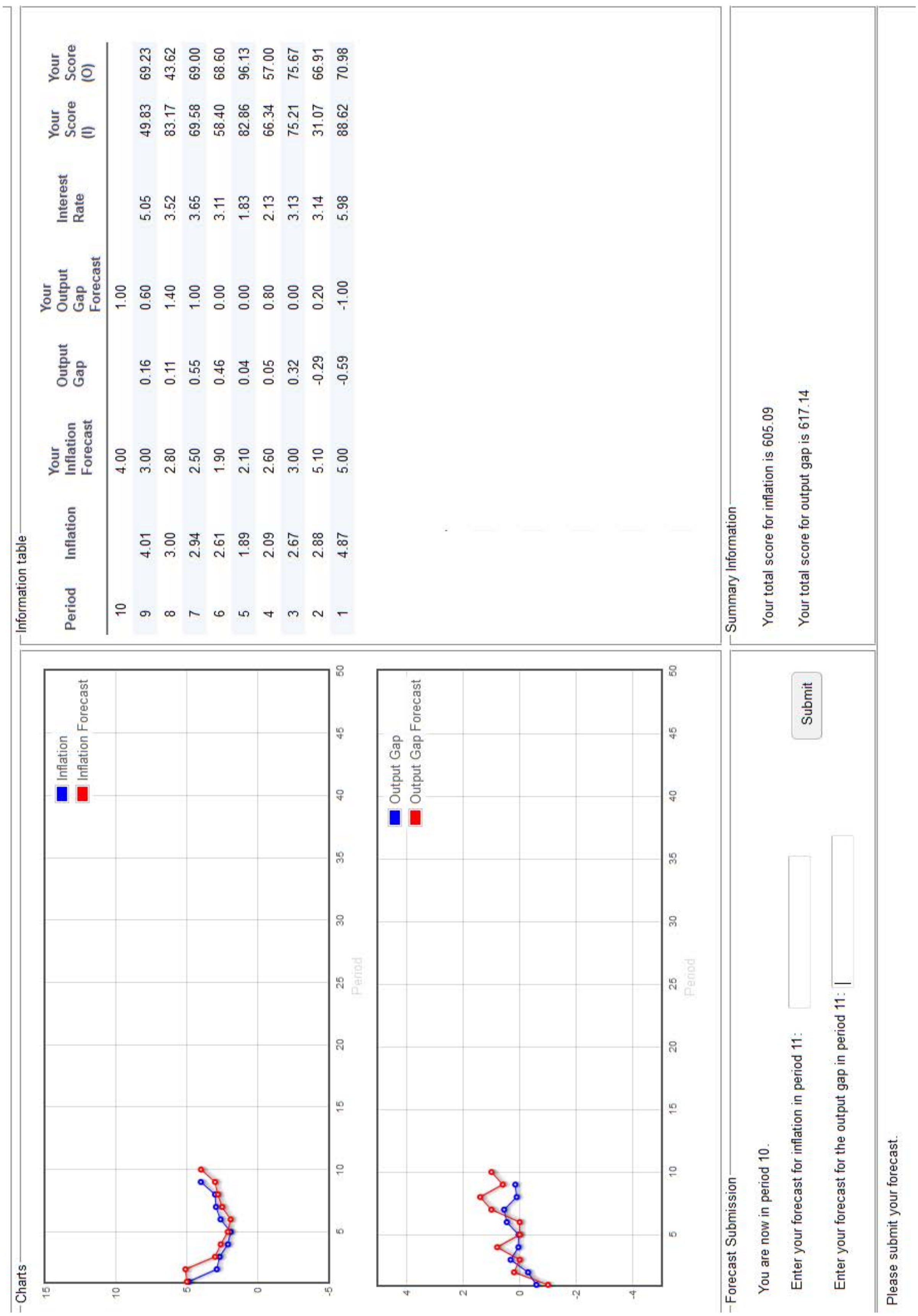


Figure 21: Screenshot