WORKING PAPERS

N° TSE - 963

November 2018

"Managing Self-organization of Expectations through Monetary Policy: a Macro Experiment"

T. Assenza, P. Heemeijer, C.H. Hommes and D. Massaro

Toulouse School of Economics

Managing Self-organization of Expectations through Monetary Policy: a Macro Experiment

T. Assenza[a,b], P. Heemeijer[c], C.H. Hommes[d,e,*], D. Massaro[f,b]

[a] *Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France*
[b] *Complexity Lab in Economics*
[c] *Symetrics*
[d] *CeNDEF, University of Amsterdam*
[e] *Tinbergen Institute, Amsterdam*
[f] *Department of Economics and Finance, Università Cattolica del Sacro Cuore, Milan*

Abstract

The New Keynesian theory of inflation determination is tested in this paper by means of laboratory experiments. We find that the Taylor principle is a necessary condition to ensure convergence to the inflation target, but it is not sufficient. Using a behavioral model of expectation formation, we show how heterogeneous expectations tend to self-organize on different forecasting strategies depending on monetary policy. Finally, we link the central bank ability to control inflation to the impact that monetary policy has on the type of feedback –positive or negative– between expectations and realizations of aggregate variables and in turn on the composition of subjects with respect to the type of forecasting rules they use.

Keywords: Laboratory Experiments, Monetary Policy, Expectations, Taylor principle.
JEL: C91, C92, D84, E52.

---

1. Introduction

The recent literature on inflation dynamics has questioned the ability of the "Taylor principle" to uniquely pin down the inflation path in the baseline rational expectations (RE) New Keynesian (NK) model (see Cochrane (2011) among others). The aim of the present paper is to shed new light on this debate by means of laboratory experiments and to empirically test for the effectiveness of the Taylor principle as a device to pin down

*Corresponding author: C.H.Hommes@uva.nl, +31(0)205254246, Valckenierstraat 65, 1018 XE Amsterdam.

inflation. The advantage of an experimental approach is that no a-priori assumption needs to be placed on agents' beliefs. Instead, expectations are directly elicited from incentivized human subjects participating in the experiment.

In NK models under rational expectations, inflation control is obtained through monetary policy satisfying the "Taylor principle" (see e.g. Woodford, 2003). When the nominal interest rate reacts more than one-for-one to deviations of inflation from its target, there exists a unique non-explosive equilibrium path, also labeled as "forward-stable" (FS) solution (García-Schmidt and Woodford, 2015). The FS-RE solution is then typically selected as the one determining inflation dynamics in the model.

Cochrane (2011), however, shows that there exist other RE solutions that cannot be ruled out by any transversality condition or economic principle. Although the Taylor principle holds, these "non-fundamental" (NF) solutions (Evans and McGough, 2018) are explosive and satisfy all relevant equilibrium conditions. The existence of NF-RE and the ability of the Taylor principle to pin down uniquely inflation dynamics are at the root of the debate on inflation control, surveyed in Section 2. Given the strong linkage, in the NK framework, between inflation dynamics and inflation expectations, the focus has shifted on the ability of central banks to manage expectations via Taylor rules. The literature has then investigated the role of expectation formation in shaping inflation dynamics by considering mild departures from RE (see e.g. McCallum, 2009; García-Schmidt and Woodford, 2015; Farhi and Werning, 2017; Gabaix, 2018; Evans and McGough, 2018; Mankiw and Reis, 2002; Coibion and Gorodnichenko, 2015; Angeletos and Lian, 2018, among others) In this paper we do not impose a-priori the type of expectations, and let them be directly elicited from participants in the experiment. Therefore, an advantage of our approach is that we can study the Taylor principle without taking a stand on the form of expectations.

In our experiment subjects are asked to forecast inflation and the output gap in an artificial NK economy and their rewards depend solely on the accuracy of these forecasts. Forecasts are then aggregated and used as inputs into a computerized NK model, which describes realizations of inflation and the output gap as functions of such forecasts and exogenous disturbances.[1] This process then repeats itself for a fixed number of periods. Our experimental economic systems are therefore "self-referential" (Marcet and Sargent, 1989) in the sense that expectations affect the data-generating process, which in turn affects expectations. As noted by Eusepi and Preston (2018), expectation errors in such environments, characterized by a dynamic feedback between expectations and realizations of aggregate variables, may propagate through the system, becoming self-fulfilling and causing instability. We use this setup to investigate whether the FS-RE solution emerges as the equilibrium outcome in the experimental economies under different monetary policy regimes by considering different parameterizations of a Taylor-type interest rate rule.

Our contribution is threefold. First, we establish that Cochrane's results on multiplicity of equilibria, do not only emerge in rational or near rational expectations settings. We also find them in a set up in which expectations are elicited from human subjects participating in the experiment. In other words, we reinforce Cochrane's results finding that the Taylor principle is a necessary, but not sufficient condition for stability and uniqueness of the equilibrium path of inflation. Second, we revisit Cochrane's results and reframe them in terms of positive versus negative expectation feedbacks. In particular, we show that the conditions for the emergence of a FS-RE solution relate to the existence of strong enough negative feedbacks.[2] Third, we show that in a heterogeneous expectations setting the convergence to a stable equilibrium is driven by a composition effect. More precisely, convergence to a stable equilibrium obtains when the share of agents adopting an adaptive expectation rule is large enough. A direct policy implication of this result is that the central bank can actually achieve convergence by managing the share of agents using a specific expectation rule. We show that this can be implemented by manipulating the relative size of the negative feedback by tuning the reaction of the policy rule to deviations of inflation from its target. In other words, the central bank can manage the composition of expectation rules adopted by agents,

[1]Aggregate outcomes computed in our laboratory economies are consistent with the notion of "temporary equilibria" in the sense that they result from first-order conditions of (computerized) households and firms given subjects' forecasts (see e.g. García-Schmidt and Woodford, 2015; Farhi and Werning, 2017; Eusepi and Preston, 2018).

[2]Negative (positive) expectations feedback means that the average forecast has a negative (positive) effect on the realized aggregate variable.

and achieve convergence to the target, by implementing an aggressive monetary policy that in turn increases the "size" of the negative feedback.

The paper is organized as follows. Section 2 relates our work to the existing literature, presents the theoretical framework and describes different monetary policy regimes. Section 3 describes the design of the experiment and shows the experimental results. Section 4 presents the model used to explain self-organization of individual expectations and the emergence of aggregate behaviors observed in the experiment. This section also discusses how the central bank can influence this process through monetary policy in order to achieve convergence to the target equilibrium. Section 5 concludes.

2. Related literature

The aim of this section is twofold. First, we describe the theoretical framework that we use in the experiment and second, we place it in the debate about inflation control via Taylor rules within the NK model.

In the following we adopt the standard New Keynesian workhorse model described by[3]

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

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

$$i_t = \text{Max}\{\bar{\pi} + \gamma + \phi_\pi(\pi_t - \bar{\pi}), 0\}\,. \quad (3)$$

Eq. (1) is the dynamic IS curve, Eq. (2) is the New Keynesian Phillips curve (NKPC) and Eq. (3) is the monetary policy rule, with a zero lower bound (ZLB), implemented by the monetary authority in order to keep inflation at its target value $\bar{\pi}$. Variables y_t and $\bar{y}^e_{t+1}$ denote respectively the actual and average expected output gap, i_t is the nominal interest rate, π_t and $\bar{\pi}^e_{t+1}$ denote respectively the actual and average expected inflation rates, $\bar{\pi}$ is the inflation target. Parameter φ is the intertemporal elasticity of substitution of consumption,

[3]Micro-founded NK models consistent with heterogeneous expectations have been derived by Branch and McGough (2009), Kurz et al. (2013), Massaro (2013) and Woodford (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 et al. (2013) in which deviations of average agents' forecasts of individual future consumption (prices) from average forecast of aggregate consumption (price) enter the error terms.

λ denotes the slope of the NKPC, ρ is the discount factor, γ is the natural interest rate. The coefficient ϕ_π measures the response of the nominal interest rate i_t to deviations of the inflation rate π_t from its target $\bar{\pi}$. Finally and g_t and u_t are exogenous disturbances, which can be thought of a demand shock and a cost push shock respectively. When the ZLB is not binding, by substituting for the monetary policy rule in Eq. (3), the model (1) – (3) can be reduced to a two variables system and written in matrix form as:

$$z_t = \mathrm{A} + \mathrm{M}\,\bar{z}^e_{t+1} + \mathrm{C}\,\epsilon_t\ , \tag{4}$$

where $z = (y, \pi)'$ is the vector of endogenous variables, $\bar{z}^e = (\bar{y}^e, \bar{\pi}^e)'$ is the vector of average forecasts and $\epsilon = (g, u)'$ is the vector of exogenous disturbances.[4] When expectations are rational and the Taylor principle is satisfied ($\phi_\pi > 1$), the model admits a FS-RE solution of the form:

$$z_t = \Theta^{FS} + \mathrm{C}\epsilon_t\ , \tag{5}$$

with $\Theta^{FS} = (I - \mathrm{M})^{-1}\mathrm{A}$, while the form of matrix C depends on the assumptions placed on the observability of the shocks. However, Cochrane (2011) argues that, in the context of NK models, the Taylor principle does not eliminate equilibrium indeterminacy. In particular there exists a NF-RE solution of the form:

$$z_t = \Theta^{NF} + \Phi^{NF} z_{t-1} + \mathrm{C}\epsilon_t\ , \tag{6}$$

with $\Theta^{NF} = (-\mathrm{M})^{-1}\mathrm{A}$, and $\Phi^{NF} = \mathrm{M}^{-1}$, while the form of matrix C depends on the assumptions placed on the observability of the shocks.

McCallum (2009) argues that a necessary condition for a RE equilibrium to be considered as representative of aggregate behavior in actual economies, is that agents should be able to learn this equilibrium from data generated by the economy itself. On these grounds, McCal-

[4] Coefficient matrices A, M and C are defined as follows:

$$\mathrm{A} \equiv \begin{pmatrix} \frac{\varphi\bar{\pi}(\phi_\pi-1)}{1+\lambda\varphi\phi_\pi} \\ \frac{\lambda\varphi\bar{\pi}(\phi_\pi-1)}{1+\lambda\varphi\phi_\pi} \end{pmatrix},\ \mathrm{M} \equiv \begin{pmatrix} \frac{1}{1+\lambda\varphi\phi_\pi} & \frac{\varphi(1-\phi_\pi\rho)}{1+\lambda\varphi\phi_\pi} \\ \frac{\lambda}{1+\lambda\varphi\phi_\pi} & \frac{\lambda\varphi+\rho}{1+\lambda\varphi\phi_\pi} \end{pmatrix},\ \mathrm{C} \equiv \begin{pmatrix} \frac{1}{1+\lambda\varphi\phi_\pi} & \frac{-\varphi\phi_\pi}{1+\lambda\varphi\phi_\pi} \\ \frac{\lambda}{1+\lambda\varphi\phi_\pi} & \frac{1}{1+\lambda\varphi\phi_\pi} \end{pmatrix}.$$

lum proposes "least-squares learnability" as an equilibrium selection device and shows that, when the Taylor principle is satisfied, the NK model with least-squares learning converges to the FS-RE equilibrium. Cochrane (2009) objects to the results derived in McCallum (2009) on the grounds that they hinge on observability of contemporaneous exogenous shocks. This assumption is indeed hard to defend for the relevant exogenous shocks in the NK model, e.g. aggregate productivity, preference or monetary policy shocks. Evans and McGough (2018) extend the results of McCallum (2009) to the case of unobservable shocks. In this case NF-RE solutions are never learnable, while the FS-RE equilibrium is learnable provided that the positive feedback from expectations to realizations of the endogenous variable being forecast is not too large, as in the case of a NK model satisfying the Taylor principle.

Our paper is directly related to this debate. In particular, our evaluation of the effectiveness of the Taylor principle for inflation determinacy is consistent with the principle put forward in McCallum (2009) and Evans and McGough (2018): subjects have imperfect information about the exact functioning of the economy they are participating in, but they can nevertheless learn the RE equilibrium through properly designed monetary policy. There are however some important differences. The first obvious difference with the least-squares learnability approach is that we do not postulate any learning mechanism, having instead real human subjects learning in the experimental economies. The second difference concerns the information set available to learning agents. In fact, contrarily to McCallum (2009) and Evans and McGough (2018), contemporaneous realizations of aggregate variables are not available to subjects when forecasting future inflation and output gap.[5] This assumption addresses the simultaneity issue raised by Cochrane (2009), i.e. how to interpret an equilibrium in which agents are forecasting based on the same endogenous variables being determined. Our conclusions regarding the effectiveness of the Taylor principle differ from those obtained under least-squares learning since we find that the Taylor principle is not a sufficient condition to ensure convergence to the FS-RE equilibrium.

Given the strong linkage in the NKPC between expectations and inflation dynamics,

[5] In our experimental implementation we consider unobservable IID exogenous disturbances with zero mean. Moreover, since realizations of endogenous variables z_t in period t depend on expectations z^e_{t+1} formed in period t, subjects in the experiment do not observe contemporaneous variables when making forecasts.

the role of beliefs formation has been widely investigated. García-Schmidt and Woodford (2015) have developed the concept of "reflective equilibrium". In particular, they posit a continuous belief revision process in which, given a conjecture about average forecasts, agents refine expectations using their knowledge of the structural equations governing the economy. In this framework issues of indeterminacy are sidestepped as, for a given level of reflection, the equilibrium outcome is unique. Moreover, when the Taylor principle is satisfied, the dynamics of the NK model under the reflective process converge to the FS-RE solution as the degree of reflection increases. Farhi and Werning (2017) adopt a form of bounded rationality based on a discrete deductive procedure rather than continuous, known as "level-k thinking" (see Nagel, 1995). Within the context of a NK model with incomplete markets, they show that the level-k equilibrium converges to the RE with complete markets as k increases only when the Taylor principle is satisfied. The main difference between our approach and both the "reflective" and "level-k thinking" is that the latter assume an iterative reasoning based on knowledge by agents of the correct quantitative specification of the economic structure, while our subjects have imperfect structural knowledge of the economy. Our experimental results show that, even without full information, monetary policy can ensure coordination on the FS-RE equilibrium.

Gabaix (2018) introduces partially myopic agents and shows that, if bounded rationality is strong enough, the NK model exhibits a unique bounded equilibrium even without the Taylor principle. Angeletos and Lian (2018) study the effect of monetary policy focusing on the forward guidance puzzle in a NK model with full rationality and informational frictions, showing how the absence of common knowledge may rationalize the kind of myopia postulated in Gabaix (2018). Mankiw and Reis (2002) propose a framework in which agents receive perfect information infrequently due to slow diffusion of information. In a framework with imperfectly informed firms, Barrdear (2018) shows that a unique bounded equilibrium emerges in the NK model regardless of whether the Taylor principle is satisfied. Our experimental findings show instead that the Taylor principle is a necessary, though not sufficient, condition to observe convergence to the FS-RE equilibrium. In this paper, contrary to this literature, we do not posit a priori a specific form of expectations, instead we rely on laboratory experiments to elicit them (see Section 3). By doing so we do not restrict ourself to a particular beliefs theory. In this respect our paper relates to the literature on macro experiments in controlled laboratory environments, (see Duffy, 2016, for a recent overview). Our experiment is a Learning-to-Forecast Experiment (LtFE), a design first proposed by Marimon and Sunder (1993) to study expectations dynamics in the laboratory. In recent years a number of LtFEs have been conducted within the NK framework to investigate inflation persistence (Adam, 2007), disinflationary policies (Cornand and M'baye, 2016), the importance of the expectation channel for macroeconomic stabilization (Kryvtsov and Petersen, 2013), and monetary and fiscal policy design at the zero lower bound (Arifovic and Petersen, 2017; Hommes et al., 2018) among other topics. Most closely related to our paper is Pfajfar and Žakelj (2018), who study the stabilization effects of different monetary policy rules by means of LtFEs. Pfajfar and Žakelj (2018) compare inflation variability under contemporaneous vs. forward-looking interest rate rules all satisfying the Taylor principle, finding that the former produces lower inflation variability. We focus instead on different contemporaneous interest rate rules, assessing the role of the Taylor principle for inflation control. Another important difference concerns the experimental design. While in Pfajfar and Žakelj (2018) participants forecast inflation only, we allow subjects to forecast both inflation and the output gap, in accordance with the theoretical NK model. By doing so we do not need to make any specific assumption on output gap expectations (as in Pfajfar and Žakelj, 2018), thus making sure that our results do not hinge on specific hypotheses placed on the belief function for the output gap.

Finally, due to the nature of the strategic environment in the NK model, our paper relates to the literature that has studied the role of strategic interactions in shaping aggregate dynamics. Fehr and Tyran (2008) show by means of laboratory experiments that aggregate behavior depends upon the strategic environment. More specifically, strategic complementarity leads to large deviations from the aggregate predictions of RE models, while strategic substitutability generate outcomes consistent with RE predictions. Our experimental environment is more complex than simple univariate systems as it is characterized by two endogenous variables, inflation and the output gap, and a policy variable, the interest rate. Depending on the policy reaction to inflation fluctuations, the system may exhibit purely positive feedbacks or a mixture of positive and negative feedbacks. In earlier LtFEs, Bao et al. (2012) have shown that, within a simple univariate environment with imperfect information, the type of expectations feedback is crucial for convergence to the RE equilibrium. In particular, negative feedback experimental markets are rather stable and converge quickly to the unique RE steady state. In contrast, positive feedback markets are rather unstable and typically do not converge, but fluctuate persistently around the RE steady state. A system characterized by positive feedbacks corresponds to a situation in which stated expectations are strategic complements, while a system characterized by negative feedbacks corresponds to a situation in which stated expectations are strategic substitutes. In our paper we link the central bank's ability to control inflation to the impact that monetary policy has on the type of feedback –positive or negative– between expectations and realizations of aggregate variables. Positive (negative) expectations feedback means that the average forecast has a positive (negative) effect on the realized aggregate variable.

More specifically in the context of the NK model, we can distinguish different monetary policy regimes according to *i*) whether the Taylor principle is satisfied and *ii*) the implied nature of expectations feedbacks in the economy described by Eq. (4). The IS curve in Eq. (1) implies that higher expected output gap leads to higher realized output gap. Moreover, since current inflation depends positively on current output gap, the NKPC in Eq. (2) implies that both higher expected inflation and higher expected output gap lead to higher realized inflation.[6] On the other hand, the linkage between expected future inflation and realized output depends on the monetary policy defined in Eq (3). In particular, if $\phi_\pi < 1/\rho$ then the system described by Eq. (4) exhibits purely positive feedbacks. If instead $\phi_\pi > 1/\rho$, then the system in Eq. (4) exhibits a mix of positive and negative feedbacks.[7] The only source of negative feedback in the NK is the monetary policy rule: when the nominal interest rate reacts aggressively enough to inflation, i.e. $\phi_\pi > 1/\rho$, then high (low) inflation expectations lead to a negative (positive) effect on output gap through real interest rate.[8] We can therefore

[6] Hence, the signs of $\partial y_t/\partial \bar{y}^e_{t+1}$ (M_{11} entry of M), $\partial \pi_t/\partial \bar{y}^e_{t+1}$ (M_{21} entry of M) and $\partial \pi_t/\partial \bar{\pi}^e_{t+1}$ (M_{22} entry of M) are positive and independent of monetary policy.

[7] In fact, if $\phi_\pi < 1/\rho$ then $\partial y_t/\partial \bar{\pi}^e_{t+1} > 0$ and all entries of matrix M are positive. While if $\phi_\pi > 1/\rho$, then $\partial y_t/\partial \bar{\pi}^e_{t+1} < 0$.

[8] Notice that the threshold value $1/\rho$ is larger than 1 since parameter $0 < \rho < 1$ denotes the time discount factor. We remark that a higher reaction coefficient ϕ_π also weakens the existing positive feedbacks since all positive entries of matrix M are monotonically decreasing, though rather flat, functions of ϕ_π. Given the

identify three qualitatively different policy regimes. In the first regime ($\phi_\pi \leq 1$) the Taylor principle is not satisfied and the economy exhibits purely positive feedbacks. In the second regime ($1 < \phi_\pi < 1/\rho$) the monetary policy rule satisfies the Taylor principle but the model is still characterized by purely positive feedbacks. In the third regime ($\phi_\pi > 1/\rho$) the Taylor principle is satisfied and the system presents a mix of positive and negative feedbacks. As described below, we experiment with different parameterization of the policy rule in Eq. (3) belonging to these different policy regimes.

We show that convergence to the FS-RE equilibrium depends on the strength of negative feedbacks introduced in the system by monetary policy via the effect of interest rate on aggregate demand. Interestingly, Cornand and Heinemann (2018) show that, in a NK model with RE, monetary policy affects strategic uncertainty, turning pricing decisions into strategic substitutes when the Taylor principle is satisfied.

3. Experiment

In our Learning-to-Forecast experiment subjects are asked to predict inflation and the output gap. These forecasts are then used to compute subsequent realizations according to the NK model described in Section 2, with structural parameters set as in Clarida et al. (2000), i.e. $\rho = 0.99$, $\varphi = 1$ and $\lambda = 0.3$. The inflation target is set at $\bar{\pi} = 2\%$, while the natural interest rate is set at $\gamma = 4\%$. Shock g_t and u_t are independent and normally distributed, with mean 0 and standard deviation 0.1. Before describing the experiment in more detail, we first discuss the treatments implemented in our LtFE.

3.1. Treatments

The treatments implemented in the experiment are motivated by the theoretical results on qualitatively different policy regimes described in Section 2. There are four treatments, differing only in the reaction coefficient ϕ_π of the interest rate rule describing monetary policy. By analyzing the experimental results in the four treatments we will be able to investigate both the role of the Taylor principle and of the "size" of the negative feedback

assumed parameterization, $M_{11} \in [0.77, 0.69]$, $M_{21} \in [0.23, 0.21]$, and $M_{22} \in [0.99, 0.90]$ for $\phi_\pi \in [1, 1.5]$.

in stabilizing our economy. Table 1 summarizes the different treatments implemented in the experiment.

[Insert Table 1 here]

In the first treatment $T1$ the policy rule coefficient is set to $\phi_\pi = 1$. Monetary policy in $T1$ belongs therefore to the regime in which the Taylor principle is not satisfied and the system exhibits purely positive feedbacks. With $\phi_\pi = 1$ the determinant of matrix $I - \mathrm{M}$ is zero, implying a continuum of constant stable solutions so that the FS-RE equilibrium is not unique.[9] Moreover, when $\phi_\pi = 1$, the eigenvalues $|\lambda_1| < |\lambda_2|$ of M^{-1} are such that $|\lambda_1| = 1$ and $|\lambda_2| > 1$, so that the NF-RE solution describes unstable equilibrium paths. We then consider small perturbations around the threshold case $\phi_\pi = 1/\rho$.[10] In particular, in the second treatment $T2$ the policy rule coefficient is set to $\phi_\pi = 1.005$, while in the third treatment $T3$ the reaction coefficient is set to $\phi_\pi = 1.015$. Treatments $T2$ and $T3$ implement both a policy regime in which the Taylor principle is satisfied. They, however, differ in terms of the type of feedback. In $T2$ the economy exhibits positive feedback only, while in $T3$ it shows a mix of positive and negative feedbacks. Note that by comparing the outcomes of $T1$ vs. $T2$, both characterized by purely positive feedback, we can assess whether a monetary policy rule satisfying the Taylor principle is a necessary and sufficient condition to ensure convergence (if any) to the unique FS-RE equilibrium. While, by comparing the outcomes in $T2$ vs. $T3$, characterized by purely positive feedback and a mix of positive and negative feedback respectively, we can determine whether the mere presence of negative feedbacks is enough to ensure convergence (if any) to the unique FS-RE equilibrium. Finally, the last treatment $T4$ considers the policy parameter originally proposed by Taylor, i.e. $\phi_\pi = 1.5$. Treatments $T3$ and $T4$ belong to the same policy regime, with the difference between $T3$ and $T4$ being the size of the negative feedback. The feedback from expected inflation to realized output $\partial y_t / \partial \bar{\pi}^e_{t+1}$ is a decreasing function of ϕ_π, so that the higher ϕ_π the more negative $\partial y_t / \partial \bar{\pi}^e_{t+1}$. By comparing the outcomes in $T3$ vs. $T4$ we can determine whether convergence (if any) to the FS-RE depends on the size of the negative feedback. Finally, note that, under

[9] Note that $\mathrm{Det}(I - \mathrm{M}) = \lambda\varphi(\phi_\pi - 1)/(1 + \lambda\varphi\phi_\pi)$. Given that φ, λ and ϕ_π are positive coefficients, when $\phi_\pi = 1$ then $\mathrm{Det}(I - \mathrm{M}) = 0$. On the contrary, whenever $\phi_\pi \neq 1$, matrix $I - \mathrm{M}$ is invertible.

[10] Note that $1/\rho$ is approximately 1.01 given the calibrated value of the time discount factor $\rho = 0.99$.

RE the FS and NF solutions depend solely on whether the Taylor principle is satisfied or not, while the nature of expectations feedback plays no role.

3.2. Procedures and implementation

The design of the experiment is a between-subjects design with within session randomization. At the beginning of each session, all participants are divided into groups (experimental economies) of six. Subjects only interact with other subjects in their experimental economy, without knowing who they are. Subjects are assigned the fictitious role of professional forecasters and they are asked to forecast inflation and the output gap. The average forecasts of all subjects in each economy are then used to calculate the realizations of inflation and output gap according to the NK model in Section 2. In each period t subjects make forecasts for period $t+1$. Their information set (visualized on their screen as numbers and partly also in graphs) is composed of: 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 indicating how close their past forecasts were to realized values up to period $t-1$.[11] Contemporaneous realizations of the small IID shocks are not observable. Moreover, the noise series used in the model equations differed across groups within each treatment, but the sets of noise series used in the four treatments were the same. Fig. B.7 in Appendix B displays the computer interface as visualized by the participants in the experiment.

Subjects' rewards depend on their forecasting performance. At the end of the experiment it is randomly determined whether a participant is paid for inflation or output gap forecasting. The final scores for inflation and output gap forecasting are given by the sums of the respective forecasting scores over all periods. The score of subject i for e.g. inflation forecast in period t is computed as $100/(1+|\pi^e_{i,t}-\pi_t|)$, where $\pi^e_{i,t}$ denotes subject i's forecast for period t and π_t realized inflation in period t (the score is computed in the same way for the output gap). Therefore rewards decrease with the distance of the realizations from their forecasts. In the instructions, subjects receive a qualitative description of the economy that includes an explanation of the mechanisms that govern the model equations, but they do

[11] Since the information set of subjects in each period t includes realizations up to period $t-1$, forecasts for period $t+1$ are actually two-period-ahead forecasts.

not have quantitative information on the exact values of structural parameters, nor on the inflation target $\bar{\pi}$.[12] The complete instructions can be found in Appendix A.

The experiment has been programmed in Java and conducted at the CREED laboratory at the University of Amsterdam. The experiment was conducted with 144 subjects (6 groups of 6 subjects for each of the 4 treatments). After each session, participants filled out a short questionnaire. Participants were primarily undergraduate students and the average age was slightly below 22 years. About half of the participants were female, about 60% were majoring in economics or business, and about 20% were Dutch. During the experiment, participants earned "points" according to the forecasting score mentioned above. Points were then exchanged for euros at the end of each session at an exchange rate of 0.75 euros per 100 points. The experiment lasted around 2 hours, and the average earning was about 25 euros.

3.3. Results

Fig. 1 presents an overview of the experimental results. Each line depicts realized inflation (left panels) and output gap (right panels) in a single experimental economy throughout the 50 periods of the experiment. The dashed lines refer to the constant equilibrium level $\bar{\pi}$ and $(1-\rho)\bar{\pi}/\lambda$ respectively for inflation and the output gap. Before describing the results in more detail, we note that, for practical reasons, we imposed bounds on the forecasts that subjects could input in the computer program. In particular the upper and lower bounds for both inflation and the output gap were respectively +100% and -100%, thus not very restrictive. Subjects were not informed ex-ante about these bounds and a pop-up message would appear on their screens only in case their forecasts were outside the allowed range. We interpret scenarios in which these constraints were binding as laboratory evidence of the possibility of subjects' coordination on explosive paths. The erratic behavior typically observed in experimental economies after subjects reach these bounds is not very meaningful from an economic point of view. Complete data for all groups separately including individual

[12] Given that our experiment is a two-period-ahead LtFE, after reading the instructions subjects are asked 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 they are participating in, inflation has historically been between -5% and 15% and the output gap between -5% and 5%.

forecasts can be found in Appendix C.

[Insert Fig. 1 here]

The first row of Fig. 1 displays realized inflation and output gap in treatment $T1$. Inflation and the output gap never converge to the equilibrium defined by the target $\bar{\pi}$. This is not necessarily surprising since the FS-RE is indeterminate in $T1$. In four out of six economies (groups 2, 4, 5 and 6) we observe explosive dynamics, with inflation forecasts rising to the upper bound on allowed forecasts. Reversal of the trend in these economies typically occurs when participants reach this upper bound.[13] As mentioned before, the ensuing large oscillations do not have a clear economic interpretation. We note that in these economies the output gap does not explode immediately with inflation. In fact, the impact of real interest rate on output is close to zero since $\phi_\pi = 1$. On the other hand, when the upward trend in inflation is reversed and deflationary spirals occur, the nominal interest rate hits the ZLB and the economy enters a severe recession. In one economy (group 3) we observe convergence to a non-fundamental steady state, while in another (group 1) we observe slow oscillations away from the target equilibrium.[14] The second row of Fig. 1 shows the dynamics of inflation and the output gap in treatment $T2$. Although the Taylor principle is satisfied, we only observe convergence to the unique FS-RE equilibrium in one economy out of six (group 3). All other groups do not converge to the FS-RE equilibrium. One economy (group 2) converges to an *almost self-fulfilling* stable equilibrium (see Hommes, 2013). The latter is characterized by coordination of expectations around a constant value which, although mathematically not a steady state, is hardly distinguishable from an equilibrium due to an eigenvalue very close to 1 and the presence of exogenous disturbances. Three out of six economies (groups 4, 5 and 6) display the same explosive behavior observed in treatment $T1$, while one economy (group 1) is characterized by sustained oscillatory behavior away from steady state. The third row of Fig. 1 presents aggregate dynamics in treatment $T3$. Strikingly, the mere presence of a

[13]In treatment $T1$ group 6 the upward trend in inflation is interrupted due to one participant who predicted -100% in the attempt to reverse the trend. Given that inflation rose to about 40% before this event, we consider it as evidence of explosive behavior.

[14]In treatment $T1$ group 1, a participant committed a typing error swapping inflation and output gap forecasts. This caused an interruption of the upward trend in inflation. We conjecture that, without the typing error, group 1 would have also experienced explosive dynamics.

small negative feedback eliminates coordination of subjects on unstable paths. In fact, we do not observe explosive dynamics in any of the experimental economies. Instead, all economies oscillate much closer to target when compared to treatments $T1$ and $T2$, with the exception of one economy (group 6) which stabilizes on an *almost self-fulfilling* equilibrium after about 30 periods of oscillatory behavior. Finally, the last row of Fig. 1 presents the results for treatment $T4$. The difference with all other treatments is remarkable: all experimental economies converge to the unique FS-RE equilibrium.

In what follows we investigate further differences between treatments. As argued in Section 3.1, by comparing the outcomes of $T1$ vs. $T2$ we can test whether the Taylor principle is a necessary and sufficient condition to ensure convergence to the unique FS-RE equilibrium. To this end, we compute the mean squared deviations (MSE) of inflation and the output gap from the target equilibrium in both $T1$ and $T2$ and perform a Wilcoxon rank-sum test.[15] According to the standard NK theory on inflation control, one would expect a significant difference between the two treatments, since monetary policy in $T2$ does satisfy the Taylor principle. The test does not reject the null that MSE in $T1$ is equal to MSE in $T2$ for both inflation and the output gap (p-values equal to 0.47 and 0.65 respectively), confirming the graphical evidence presented in Fig. 1 that the Taylor principle is not a sufficient condition for convergence to the FS-RE equilibrium.[16] We then compare experimental outcomes in $T2$ vs. $T3$ to assess whether by simply adding small negative feedbacks in the system, monetary policy can ensure convergence to the target. The Wilcoxon rank-sum test rejects the null of equal MSE for the output gap in $T2$ and $T3$ (p-value equal to 0.01), though the result is not as clear-cut for inflation (p-value equal to 0.06). Although aggregate dynamics are much closer to target in $T3$ when compared to $T1$ and $T2$, the presence of negative feedbacks in the system is not a sufficient condition for convergence to the FS-RE equilibrium. The Wilcoxon signed-rank test rejects the null that average inflation in $T3$ is equal to target (p-value equal to 0.03), while it does not reject it for the output gap (p-value equal to 0.09).[17] Finally, we compare

[15]In all treatments' comparisons we allow for an initial learning phase and consider data starting from period 15.

[16]Strictly speaking, the Wilcoxon rank-sum test tests the null-hypothesis that the distribution does not change against the alternative that it shifts between treatments.

[17]Technically, the Wilcoxon signed-rank test tests the null hypothesis that the distribution of average

$T3$ vs. $T4$ to verify whether convergence to the FS-RE depends on the strength of negative feedbacks. Realizations of aggregate variables in treatment $T4$ are clearly centered around the FS-RE equilibrium. This is largely confirmed by a Wilcoxon signed-rank test (p-values equal to 0.44 and 0.69 respectively for inflation and the output gap). We therefore conclude that, for the FS-RE equilibrium to emerge as the unique outcome, not only monetary policy has to satisfy the Taylor principle, but the negative feedback introduced in the system by the interest rate rule has to be strong enough. Moreover, the Wilcoxon rank-sum test rejects the null of equal MSE for inflation in $T3$ and $T4$ (p-value 0.001), while it does not reject it for the output gap (p-value 0.15).

4. Monetary policy and self-organization of expectations

The experimental economies presented in Section 3.3 show different types of aggregate behavior, namely explosive dynamics, persistent oscillations and convergence to (some) equilibrium. The goal of this section is to characterize individual forecasting behavior using a simple behavioral model of learning and explain the emergence of different aggregate patterns depending on monetary policy.

4.1. Heuristics switching model of expectation formation

The fact that different types of aggregate behavior arise in our experiments, both within and between treatments, suggests that heterogeneous expectations play an important role in determining aggregate outcomes. A first result emerging from the analysis of individual forecasts is that subjects tend to coordinate on a common prediction strategy, although participants in different groups may coordinate on different strategies. Coordination is, however, not perfect and heterogeneity in individual forecasts within groups persists throughout the experiment (see Appendix C). Another interesting result that emerges from experimental data is that individual forecasting behavior entails a learning process taking the form of switching from one prediction strategy to another (see Appendix D).

In light of this empirical evidence we use a heuristics switching model (HSM), which features evolutionary selection among different forecasting strategies, to characterize expec-

inflation or output gap is centered around the target.

tations dynamics and explain emergent aggregate behavior. Denoting by $\mathcal{H}$ a set of H forecasting heuristics for variable x, aggregate expectations in each period t are given by a weighted average of the forecasts resulting from these heuristics. In the context of the NK model x denotes either inflation or the output gap. The key ingredient of the model is that the weight of each heuristic $h \in \mathcal{H}$ evolves over time as a function of past forecasting performance. In particular the measure of past performance of heuristic h denoted as U_h is defined as

$$U_{h,t-1} = F(x_{t-1} - x^e_{h,t-1}) + \eta U_{h,t-2} \; , \tag{7}$$

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 attached to past errors of heuristic h. Performance uniquely depends on the most recent forecasting error when $\eta = 0$, while it is determined by all past prediction errors with exponentially declining weights when $0 < \eta < 1$, or equal weights when $\eta = 1$. Given the performance measure in Eq. (7), the weight attached to each heuristic h at time t is defined as

$$n_{h,t} = \delta n_{h,t-1} + (1-\delta)\frac{\exp(\beta U_{h,t-1})}{Z_{t-1}} \; , \tag{8}$$

where $Z_{t-1} = \sum_{h=1}^{H} \exp(\beta U_{h,t-1})$ is a normalization factor. Parameter $0 \leq \delta \leq 1$ describes inertia in the evolution of weights, while parameter $\beta \geq 0$ represents the intensity of choice, measuring the sensitivity to differences in heuristics performances. The model described by Eqs. (7)–(8) has been developed by Anufriev and Hommes (2012), along the lines of Brock and Hommes (1997), to explain different types of aggregate behavior as well as individual expectations in the asset pricing LtFE of Hommes et al. (2005).[18] The model is also related to reinforcement learning models developed in game-theoretical frameworks, (see e.g. Camerer and Ho, 1999), and to rational inattention models, (see e.g. Matějka and McKay, 2015).

In order to use the HSM for policy analysis, specific assumptions have to be made about

[18] In the original approach of Brock and Hommes (1997) the individual heuristics' choice in each period is random, with probability of selecting predictor h given by Eq. (8) with $\delta = 0$. With a continuum of agents and independent decisions Eq. (8) gives the proportion of agents using heuristics h. Given that each experimental economy consists of a small number of subjects, we interpret the weights in Eq. (8) as the weights attributed by subjects to different forecasting rules.

the forecast error function F and the types of forecasting heuristics to include in set $\mathcal{H}$. In our implementation of the model we use the same forecast error function used to incentivize subjects in the experiment, i.e. $F(x - x^e) = 100/(1 + |x^e - x|)$. Moreover, we discipline the choice of the set of heuristics $\mathcal{H}$ using experimental data. In particular, we consider heuristics describing qualitatively different types of forecasting behavior emerging from data on individual predictions. We restrict our attention to a set of four heuristics described in Table 2. Details on the analysis of individual forecasts time series are given in Appendix E.

[Insert Table 2 here]

The parameterization of the heuristics in Table 2 follows Anufriev and Hommes (2012) and it is consistent with estimated values using our experimental data (see Appendix E). Based upon the calibration in their paper, we set the model parameters $\beta = 0.4$, $\eta = 0.7$, $\delta = 0.9$.[19] We adopt therefore the same 4-type HSM that has successfully been used by Anufriev and Hommes (2012) to explain different price patterns emerged in the asset pricing experiment of Hommes et al. (2005). This illustrates the robustness of the HSM across different experimental settings.

As shown in Appendix F, different homogeneous expectations models, i.e. economies where all subjects use one of the forecasting heuristics in Table 2 to predict inflation and the output gap, can explain different observed patterns in aggregate variables. For example, coordination on forecasting rules strongly extrapolating past trends (STR) leads to explosive dynamics under all considered policy regimes, while coordination on adaptive rules (ADA) has a stabilizing effect under all considered policy regimes. However, homogeneous expectations models do not answer the question *why* coordination on certain prediction strategies emerge under different policy regimes. Our goal is to explain why subjects coordinate on a certain forecasting rule depending on monetary policy and how this leads to the emergence of different aggregate behavior.

[19] We remark that the model is not very sensitive to these parameter values (see also Anufriev and Hommes, 2012), and for different choices of the coefficients of the four heuristics in Table 2 we obtain similar results to those presented in Section 4.2.

4.2. Self-organization of heterogeneous expectations

In this section we discuss the performance of the HSM in describing experimental results and illustrate how the model explains the emergence of different aggregate behaviors. For each group, we compute one-step-ahead predictions of the HSM described in Section 4.1, and then compare them with experimental outcomes. Simulations are initialized using the first two realizations for inflation and the output gap, i.e. $\{\pi_1, y_1\}$ and $\{\pi_2, y_2\}$, with equal initial weights $n_h = 1/4$ for all heuristics. Using equal weights for periods 3 and 4 and the heuristics forecasts, we compute $\{\pi_3, y_3\}$ and $\{\pi_4, y_4\}$. Starting from period 5 dynamics are well defined and HSM forecasts are obtained using the same information available to subjects in the experiment. Table 3 reports the mean squared prediction errors averaged across groups in each treatment. We remark that simulations were truncated whenever bounds on individual predictions were reached or subjects tried to strategically reverse explosive trends.[20]

[Insert Table 3 here]

The results show that the HSM is a better predictor than any of the four heuristics alone in almost all cases. The only exceptions are the unstable economies in $T1$ and $T2$ in which the strong trend-following rule performs better in predicting the explosive behavior of aggregate variables. In fact, although the HSM encompasses the STR prediction strategy, the weights of the four rules are updated with some inertia due to a positive δ and a finite intensity of choice β. This result suggests that in situation of high instability, subjects coordinate faster, i.e. $\delta \to 0$ and $\beta \to \infty$, on forecasting rules that strongly extrapolate observed trends. The relatively high MSE registered for all models regarding output gap expectations in $T2$ is due to predictions of one participants in group 5 which, before hitting the upper bound in period 9, were consistently above the average of all other predictions (almost four times higher on average). Removing this one subject from the sample yields much lower MSE values but it does not change the models' ranking in terms of predicting power.

Figs. 2–4 illustrate how the HSM explains the emergence of different aggregate behaviors

[20] In particular, groups 2, 4, 5, and 6 in $T1$ were simulated respectively until periods 19, 25, 22, and 18, while groups 1, 4, 5, and 6 in $T2$ were simulated respectively until periods 11, 11, 9, and 21. Removing these groups from the sample does not change our qualitative results, though the level of MSE in $T1$ and $T2$ is obviously much lower when unstable economies are not considered in the analysis.

observed in the experiment, namely explosive dynamics, persistent oscillations and convergence to (some) equilibrium. Fig. 2 refers to group 4 in $T1$ as an example of explosive dynamics, Fig. 3 refers to group 5 in $T3$ as an example of persistent oscillations, while Fig. 4 refers to group 2 in $T4$ as an example of convergence to equilibrium. Results for other economies displaying the same type of aggregate behavior are qualitatively similar (see Figs. G.23–G.34 in Appendix G reporting results for all experimental economies). Left panels in Figs. 2–4 display experimental data together with the one-step-ahead predictions under the HSM. Overall, the one-step-ahed forecasts closely track experimental data and the model is able to reproduce qualitatively all different types of aggregate behavior.[21] Right panels in Figs. 2–4 depict the evolution over time of the weights of the four considered heuristics. In different groups different heuristics gain more weight after starting from a uniform distribution. In fact, the evolutionary learning process described by the HSM self-organizes into coordination on one of the four rules, which then determine (long-run) aggregate behavior.

[**Insert Fig. 2 here**]

In treatment $T1$ group 4 (Fig. 2) inflation follows an upward trend in the early stage of the experiment, triggering increasing coordination on trend-following behavior. The increasing trend in inflation is amplified by coordination on the STR forecasting rule, whose weight reaches about 90% by the end of the simulation. As noted in Section 3.3, the output gap does not explode immediately with inflation because the impact of real interest rate on output is close to zero when $\phi_\pi = 1$. Therefore, as long as the output gap remains stable, the weights of the four heuristics are similar. However, the sharp increase of the output gap towards the end of the considered time period, caused by rising inflation expectations, leads to increasing coordination on the STR rule. The emergence of explosive dynamics is thus explained by coordination of individual expectations on forecasting strategies that strongly extrapolate trends observed in the data. This behavior is consistent with the theoretical benchmark derived under homogeneous STR expectations in $T1$, i.e. explosive dynamics due to real eigenvalues outside the unit circle (see Appendix F for details).

[21] We also test for the null hypothesis of equality between observed and simulated mean and standard deviation of inflation and output gap using a Wilcoxon rank-sum test. In all cases we never reject the null using a 5% significance level.

[Insert Fig. 3 here]

In treatment $T3$ group 5 (Fig. 3) aggregate dynamics are characterized by persistent oscillations. The HSM explains sustained oscillatory behavior by coordination of most agents on a LAA rule. The observed trends in inflation and the output gap in the beginning of the experiment cause an initial coordination on trend-following behavior. However, reversal of the trend favors the LAA rule in the evolutionary competition among heuristics. In fact, in the presence of cyclical oscillations, the purely extrapolative rules WTF and STF tend to overshoot the trend reversal. On the other hand, the LAA rule uses an anchor which is given by a weighted average of the sample mean and the last observation, thus making smaller forecast errors at the turning points of the trend. For both inflation and the output gap, the LAA rule dominates reaching a peak weight of about 90% towards the end of the experiment, which slowly decreases afterwards as the amplitude of oscillations decreases in the last few periods. Oscillatory non-explosive behavior is consistent with the theoretical benchmark derived under homogeneous LAA expectations in $T3$, i.e. sustained non-explosive oscillations due to stable complex eigenvalues close to the unit circle (see Appendix F for details).

[Insert Fig. 4 here]

In treatment $T4$ group 2 (Fig. 4) dynamics converge to the FS-RE equilibrium. The initial part of the experiment is characterized by coordination on the LAA forecasting rule due to the continuous reversal of trends in aggregate variables. However, as oscillations gradually dampen, the weight of the ADA rule gradually increases. In fact, adaptive rules perform better in converging paths as they do not extrapolate past trends in observed variables. Convergence with progressively dampened oscillations is consistent with the theoretical benchmark derived under homogeneous ADA expectations in $T4$, i.e. oscillatory convergence due to complex eigenvalues within the unit circle (see Appendix F for details).

The one-step-ahead simulations show that initially heterogenous expectations tend to self-organize on common predictions strategies. A salient result is that the proportion of agents using (strong) trend extrapolation rules plays an important role for the stability of aggregate variables. Groups in which the weight of STR rules is lower are more stable than groups with a higher impact of trend following behavior. Instead, having more agents that follow adaptive expectations has a stabilizing effect on aggregate dynamics, while oscillatory behavior is associated with anchoring and adjustment heuristics.[22] In the following section we discuss how monetary policy can influence the process of self-organization of expectations, preventing coordination on destabilizing trend-following behavior and ensuring convergence to the FS-RE equilibrium.

4.3. Managing coordination on trend-following behavior through monetary policy

All experimental economies start away from, typically above, the target equilibrium.[23] By its impact on the feedback between expectations and realizations of aggregate variables, monetary policy can influence the adjustment process towards the target. In simple univariate systems, a positive feedback between expectations and realizations implies that the latter move together, so that deviations from equilibrium in one direction provide incentives to deviate in the same direction. On the contrary, in negative feedback systems, deviations of expectations from equilibrium in one direction have an impact on the endogenous variable going in the opposite direction. Previous experimental literature has shown that, in univariate markets with imperfect information, subjects are able to learn the unique RE equilibrium in negative feedback systems but do not converge to it in positive feedback systems. The NK model is different from simple univariate frameworks where the nature of expectations feedback is uniquely defined, i.e. either positive or negative. In fact, depending on monetary policy, the system can be characterized by either purely positive feedbacks or by a mix of positive and negative feedbacks. When the NK model exhibits purely positive feedbacks (treatments $T1$ and $T2$), indeterminacy arises because monetary policy is not able to correct drifts in expectations which may become self-sustaining. When the policy rule reacts aggressively enough to inflation, it introduces a negative feedback in the system (treatments

[22] Interestingly, Pfajfar and Žakelj (2018) reach a similar conclusion and note that a higher proportion of trend extrapolation increases the standard deviation of inflation while having more agents behaving according to adaptive expectations decreases the standard deviation of inflation.

[23] This is due to the fact that at the beginning of the experiment, when no realizations of aggregate variables are observed yet, forecasts tend to cluster around the midpoint of the interval of historical values given to subjects in the instructions. In the experiment the midpoints of these intervals are 5% and 0% respectively for inflation and the output gap.

$T3$ and $T4$), which has a stabilizing effect through the impact of real interest rate on aggregate demand. In order to appreciate the stabilizing effect of this negative feedback, it is instructive to look at cross-correlations, reported in Figs. 5–6, among realized and expected aggregate variables in the experiment. Note that in Figs. 5–6, the notation $\bar{\pi}^e$ and $\bar{y}^e$ refers to expectations formed in period t about inflation and the output gap in $t+1$, so that e.g. corr$(y, \bar{\pi}^e)$ refers to correlation between y_t and $\bar{\pi}^e_{t+1}$.

Fig. 5 displays cross-correlations at different leads and lags, averaged across groups, for treatments $T1$ and $T2$ characterized by purely positive feedbacks.

[Insert Fig. 5 here]

From Fig. 5(a), the first thing that one notices is that correlations are positive across the board. For example, correlation between realized inflation (output gap) and expected future inflation (output gap) is positive not only contemporaneously, but also at several leads/lags. Autocorrelations of expected inflation (output gap) are also positive for several lags. In fact, initial trends in aggregate variables are never reversed due to the absence of target rates stabilization through monetary policy. In particular, the positive correlation between the output gap and expected inflation (corr$(y, \bar{\pi}^e) > 0$) implies that there is no reduction in the output gap, via real interest rate, when inflation expectations are above target because the nominal interest rate does not react enough to inflation. Absent the correction mechanism of expectations via monetary policy, deviations from the target are either reinforced by coordination on forecasting rules that extrapolate the direction of change, hence resulting in explosive paths (see Fig. 2), or they stabilize around one of the multiple steady states. Results are very similar in treatment $T2$ as correlations in Fig. 5(b) are generally positive across variables. In fact, even if the Taylor principle is satisfied, the system exhibits purely positive feedbacks. Drifts in expectations away from the target are, in general, not corrected towards the FS-RE equilibrium and dynamics may either explode or converge to an almost self-fulfilling equilibrium.[24]

Fig. 6 shows cross-correlations for treatments $T3$ and $T4$, characterized instead by a mix of positive and negative feedbacks. We first discuss results for treatment $T4$ and then

[24] There is only one experimental economy that oscillates around the target equilibrium in $T2$.

examine treatment $T3$.

[Insert Fig. 6 here]

From Fig. 6(b), it is clear that the presence of negative feedbacks in the system significantly changes the correlation structure among variables when compared to treatments $T1$ and $T2$. As in other treatments, initial inflation expectations above target cause realized inflation to be above target as well. In this case, however, the marked reaction of the nominal interest rate causes an increase in the real interest rate so that the output gap decreases ($\mathrm{corr}(y, \bar{\pi}^e) < 0$), curbing therefore the inflationary pressure. Output gap expectations follow the decreasing trend in the output gap ($\mathrm{corr}(y, \bar{y}^e_{+1}) > 0$), further reducing inflation and subsequently inflation expectations ($\mathrm{corr}(\bar{y}^e, \bar{\pi}^e_{+1}) > 0$). Decreasing inflation and output gap expectations cause inflation to fall and eventually overshoot the target. This leads to lower real interest rate which in turn stimulates aggregate demand. This continuous trend reversal, driven by the effect of monetary policy on aggregate demand, is reflected e.g. in the observed negative autocorrelation of output gap expectations after the first lag ($\mathrm{corr}(\bar{y}^e, \bar{y}^e_i) < 0$ for $i < -1$). In this environment destabilizing trend-following strategies perform poorly, and they are driven out by stabilizing adaptive expectations in the evolutionary competition among predictors (see Fig. 4). As the weight of trend-following strategies decreases, oscillations in aggregate variables progressively dampen and the system eventually converges to the FS-RE equilibrium. In treatment $T3$ the policy reaction does introduce negative feedbacks in the system, which is reflected in the negative correlation between expected inflation and current output ($\mathrm{corr}(y, \bar{\pi}^e) < 0$) in Fig. 6(a). In fact, as in treatment $T4$, we do observe reversal of initial trends in inflation via the impact of real interest rate on aggregate demand, so that coordination on forecasting strategies that strongly extrapolate past trends is prevented (see Fig. 3). However, the impact on aggregate demand is not strong enough to quickly revert drifts in inflation expectations. In fact, although output gap expectations follow the decreasing trend in the output gap due to inflation expectations above target ($\mathrm{corr}(y, \bar{y}^e_{+1}) > 0$), their impact on realized inflation is mild, so that inflation expectations may still increase despite a negative trend in output gap expectations ($\mathrm{corr}(\bar{y}^e, \bar{\pi}^e_{+1}) < 0$). In other words, the signals that subjects receive are not strong enough to promptly correct their expectations. The sluggish dynamics observed in treatment $T3$ are reflected in the observed positive autocorrelation of e.g. output gap expectations until the third lag ($\text{corr}(\bar{y}^e, \bar{y}^e_i) > 0$ for $-4 < i < 0$).

How can monetary policy manage the self-organization process of expectations and ensure determinacy of the FS-RE equilibrium? Our results show that, in the presence of imperfect information, obeying the Taylor principle does not necessarily lead to convergence to the target. In fact, even if monetary policy reacts more than point-to-point to inflation, the NK model may still exhibit purely positive feedbacks. Results from treatment $T2$ show that in such an environment, when a majority of individuals use a trend-following strategy, other individuals have an incentive to use such strategy too, thus reinforcing trends in aggregate variables. An insight emerging from our analysis is that the introduction of negative feedbacks via monetary policy is a necessary condition to prevent coordination on trend-following behavior. However, the mere presence of negative feedbacks is not sufficient for the FS-RE to emerge as the unique outcome in the experimental economies, as shown in treatment $T3$. To ensure convergence to the desired equilibrium, monetary policy has to be aggressive enough to quickly correct drifts in expectations towards the target. How aggressive should then monetary policy be to control inflation? It is important to note that, as long as matrix M, mapping expectations into realizations of aggregate variables, is close to having an eigenvalue equal to 1, the system exhibits sluggish adjustment dynamics and it may converge to almost self-fulfilling equilibria. This is in fact the case for treatment $T3$, in which the absolute value of largest eigenvalue is about 0.98. For subjects to learn the FS-RE equilibrium from data generated by the economic system, the eigenvalues of matrix M have to be well within the unit circle. Results from treatment $T4$ suggest that a reaction coefficient $\phi_\pi = 1.5$, leading to a largest eigenvalue of about 0.83, is sufficient to ensure convergence to the target.

5. Conclusions

Laboratory experiments have been used in this paper to test the New Keynesian theory of inflation determination. Our results suggest that the Taylor principle does not ensure convergence to the inflation target. Using a behavioral model of expectation formation, we explain how different aggregate outcomes emerge out of a self-organization process of heterogenous expectations driven by their relative forecasting performance. We illustrate how monetary policy can prevent coordination on explosive non-fundamental equilibria and steer expectations towards the target. In particular, by introducing a strong enough negative feedback between expected inflation and aggregate demand, the central bank can avoid coordination on trend-following behavior and prevent expectation errors from becoming (partially) self-fulfilling.

Our experiment focuses on short-run forecasts. However, recent literature on forward guidance about future central bank actions has highlighted the importance of expectations at far horizons for inflation control. Future experiments within NK economies should also incorporate elicitation of long-run forecasts. Moreover, our study focuses on an heuristic switching model of expectation formation. Recent works have proposed several alternative models of expectations within the NK framework, see e.g. least-squares learning (Evans and McGough, 2018), sticky information (Mankiw and Reis, 2002), sparsity-based bounded rationality (Gabaix, 2018), rational inattention (Maćkowiak and Wiederholt, 2015), reflective equilibrium (García-Schmidt and Woodford, 2015), level-k thinking (Farhi and Werning, 2017), and imperfect information (Angeletos and Lian, 2018) among others.

References

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

Angeletos, G.-M., Lian, C., 2018. Forward guidance without common knowledge. American Economic Review 108 (9), 2477–2512.

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

Arifovic, J., Petersen, L., 2017. Stabilizing expectations at the zero lower bound: Experimental evidence. Journal of Economic Dynamics and Control 82, 21 – 43.

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

Barrdear, J., 2018. The calm policymaker: Imperfect common knowledge in new keynesian models. Working paper, Bank of England.

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

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

Camerer, C. F., Ho, T.-H., 1999. Experienced-weighted attraction learning in normal form games. Econometrica 67 (4), 827–874.

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

Cochrane, J. H., 2009. Can learnability save new-Keynesian models? Journal of Monetary Economics 56 (8), 1109 – 1113.

Cochrane, J. H., 2011. Determinacy and identification with Taylor rules. Journal of Political Economy 119 (3), 565–615.

Coibion, O., Gorodnichenko, Y., 2015. Information rigidity and the expectations formation process: A simple framework and new facts. American Economic Review 105 (8), 2644–78.

Cornand, C., Heinemann, F., 2018. Monetary Policy obeying the Taylor Principle Turns Prices into Strategic Substitutes. Discussion Paper 90, CRC Rationality and Competition.

Cornand, C., M'baye, C. K., 2016. Does inflation targeting matter? an experimental investigation. Macroeconomic Dynamics, 140.

Duffy, J., 2016. Macroeconomics: A Survey of Laboratory Research. In: Kagel, J. H., Roth, A. E. (Eds.), The Handbook of Experimental Economics, Volume 2. Princeton University Press, pp. 1–90.

Eusepi, S., Preston, B., 2018. The science of monetary policy: An imperfect knowledge perspective. Journal of Economic Literature 56 (1), 3–59.

Evans, G., McGough, B., 2018. Equilibrium selection, observability and backward-stable solutions. Journal of Monetary Economics 98, 1 – 10.

Farhi, E., Werning, I., 2017. Monetary policy, bounded rationality, and incomplete markets. Working Paper 23281, National Bureau of Economic Research.

Fehr, E., Tyran, J.-R., 2008. Limited rationality and strategic interaction: the impact of the strategic environment on nominal inertia. Econometrica 76 (2), 353–394.

Gabaix, X., 2018. A behavioral new keynesian model. Harvard working paper.

García-Schmidt, M., Woodford, M., 2015. Are low interest rates deflationary? a paradox of perfect-foresight analysis. Working Paper 21614, National Bureau of Economic Research.

Hommes, C., Massaro, D., Salle, I., 2018. Monetary and fiscal policy design at the zero lower bound – evidence from the lab. Economic Inquiry.

Hommes, C. H., 2013. Reflexivity, expectations feedback and almost self-fulfilling equilibria: economic theory, empirical evidence and laboratory experiments. Journal of Economic Methodology 20 (4), 406–419.

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

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

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

Maćkowiak, B., Wiederholt, M., 2015. Business cycle dynamics under rational inattention. Review of Economic Studies 82 (4), 1502–1532.

Mankiw, N. G., Reis, R., 2002. Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. The Quarterly Journal of Economics 117 (4), 1295–1328.

Marcet, A., Sargent, T., 1989. Convergence of least squares learning in environments with hidden state variables and private information. Journal of Political Economy, 1306–1322.

Marimon, R., Sunder, S., 1993. Indeterminacy of equilibria in a hyperinflationary world: Experimental evidence. Econometrica 61 (5), 1073–1107.

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

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

McCallum, B. T., 2009. Inflation determination with Taylor rules: Is new-Keynesian analysis critically flawed? Journal of Monetary Economics 56 (8), 1101–1108.

Nagel, R., 1995. Unraveling in guessing games: an experimental study. American Economic Review 85, 1313–1326.

Pfajfar, D., Žakelj, B., 2018. Inflation expectations and monetary policy design: Evidence from the laboratory. Macroeconomic Dynamics 22 (4), 1035?1075.

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

Woodford, M., 2013. Macroeconomic analysis without the rational expectations hypothesis. Annual Review of Economics 5 (1), 303–346.

Acknowledgments: We thank the editor Ricardo Reis and four anonymous reviewers for detailed comments and suggestions. We also thank Klaus Adam, Jess Benhabib, Jim Bullard, Fabrice Collard, Domenico Delli Gatti, John Duffy, George Evans, Patrick Fève, Seppo Honkapohja, Albert Marcet, Argia Sbordone, Shyam Sunder, Matthias Weber, John Williams and Mike Woodford for stimulating discussions and helpful comments. Domenico Massaro acknowledges financial support from the Ministry of Education, Universities and Research of Italy (MIUR), program *SIR* (grant n. RBSI144KWH). None of the above are responsible for errors in this paper.

Tables and figures

Table 1: Summary of policy regimes and characteristics of RE solutions in different treatments

Treatment	ϕ_π	Taylor principle	Expectations feedbacks	FS	NF
$T1$	1	No	Purely Positive	Indeterminate	Explosive
$T2$	1.005	Yes	Purely Positive	Unique	Explosive
$T3$	1.015	Yes	Mix Positive/Negative	Unique	Explosive
$T4$	1.5	Yes	Mix Positive/Negative	Unique	Explosive

Table 2: 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	$x^e_{4,t+1} = 0.5(\bar{x}_{t-1} + x_{t-1}) + (x_{t-1} - x_{t-2})$

Note: The term $\bar{x}_{t-1}$ denotes the average of all observations up to time $t-1$.

Table 3: MSE of one-step-ahead simulations for different models of expectation formation

	Treatment $T1$		Treatment $T2$		Treatment $T3$		Treatment $T4$	
	π	y	π	y	π	y	π	y
HSM	3.410	0.098	5.851	8.886	0.714	0.425	0.070	0.083
ADA	61.700	0.323	47.989	18.917	4.350	1.009	0.371	0.482
WTR	19.168	0.152	12.149	10.608	1.586	0.524	0.091	0.149
STR	1.161	0.133	3.599	4.579	2.049	0.690	0.212	0.349
LAA	58.794	0.271	30.221	12.992	2.355	0.559	0.195	0.110

Note: The MSE is computed over periods 5 to 49 in order to minimize the impacts of initial conditions on heuristics' weights and of "ending effects" in individual forecasts observed in several groups.

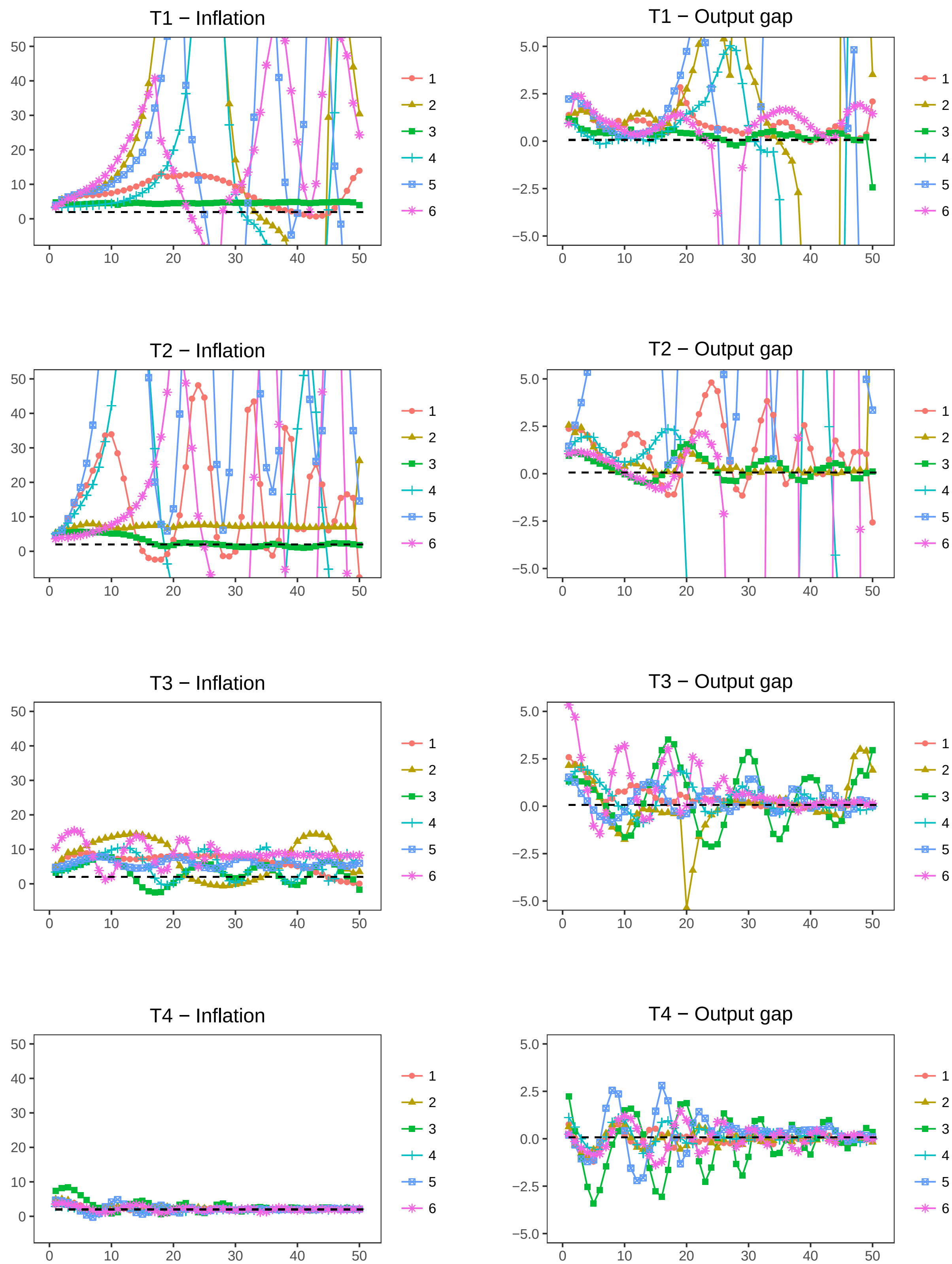


Figure 1: Inflation (left panels) and output gap (right panels) dynamics in different groups and treatments. Each line refers to one experimental economy, numbered from 1 to 6.

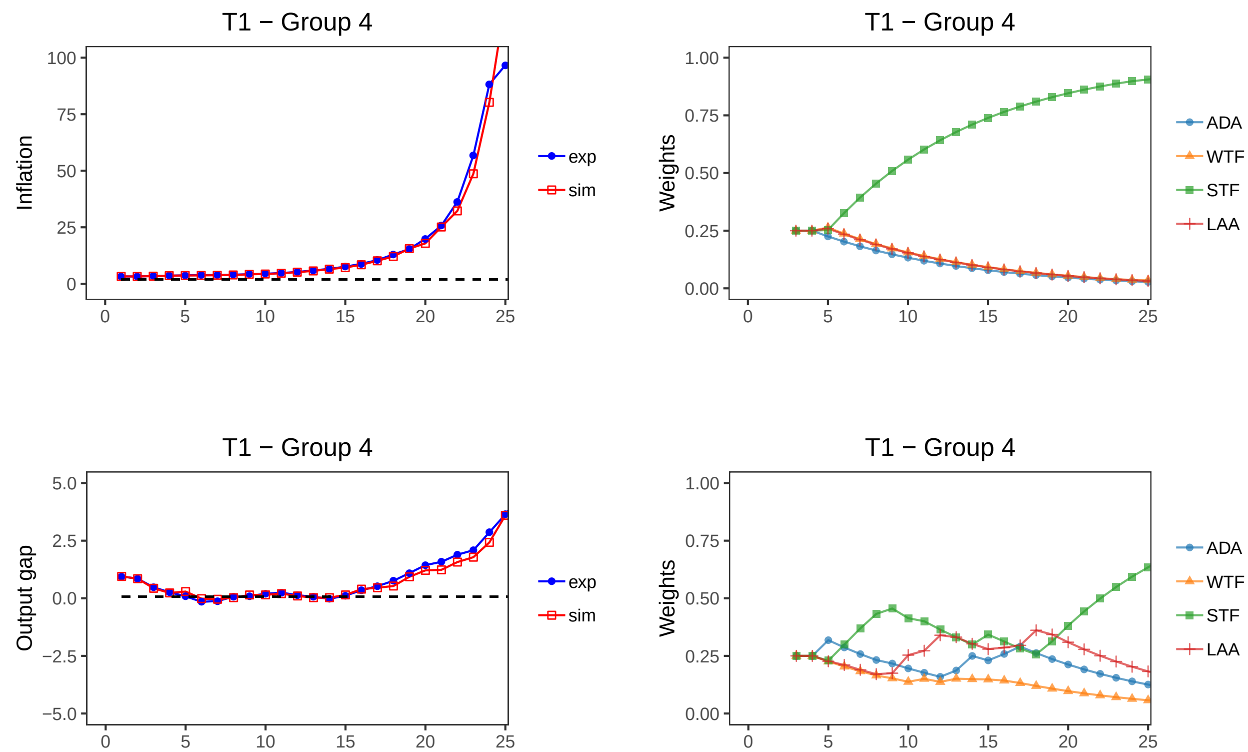


Figure 2: Realized and simulated inflation and output gap (left panels) with corresponding weights of 4 heuristics for $T1$ group 4. In the left panels, blue circles refer to experimental data while red squares refer to simulated data. In the right panels, ADA, WTR, STR and LAA refer respectively to the adaptive rule, the weak trend-following rule, the strong trend-following rule and the anchoring and adjustment rule.

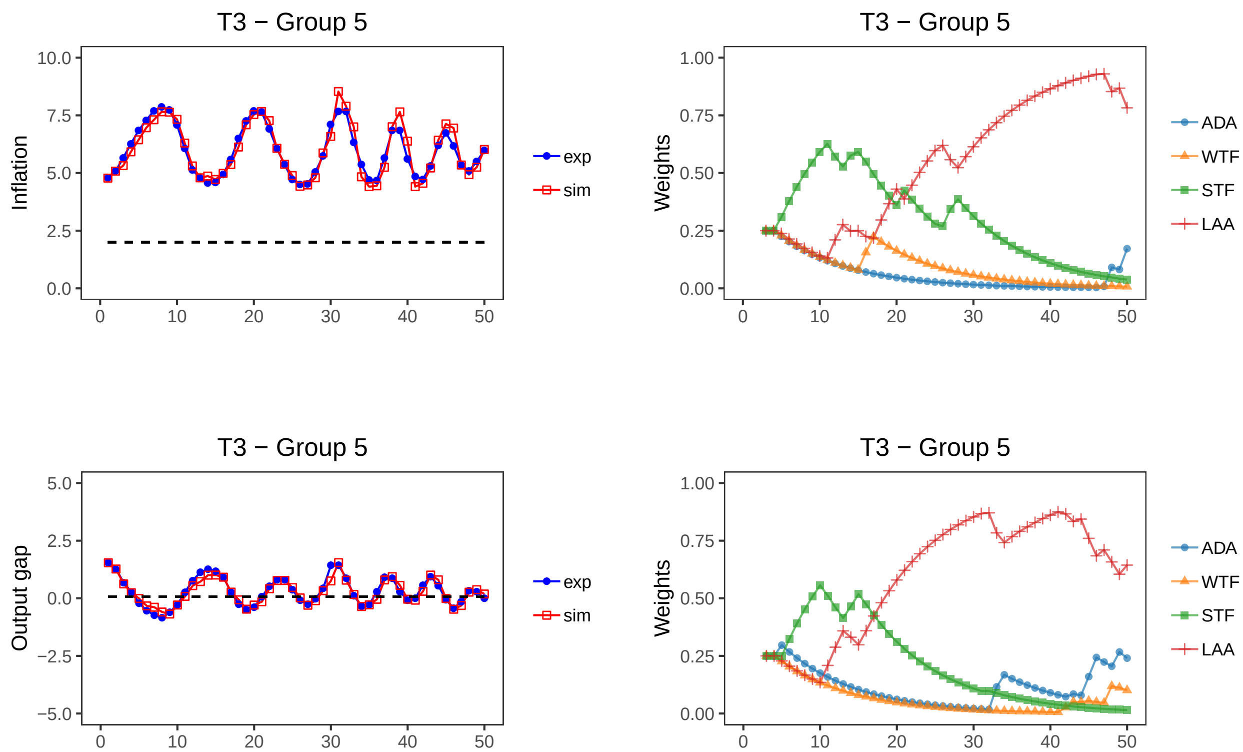


Figure 3: Realized and simulated inflation and output gap (left panels) with corresponding weights of 4 heuristics for $T3$ group 5. In the left panels, blue circles refer to experimental data while red squares refer to simulated data. In the right panels, ADA, WTR, STR and LAA refer respectively to the adaptive rule, the weak trend-following rule, the strong trend-following rule and the anchoring and adjustment rule.

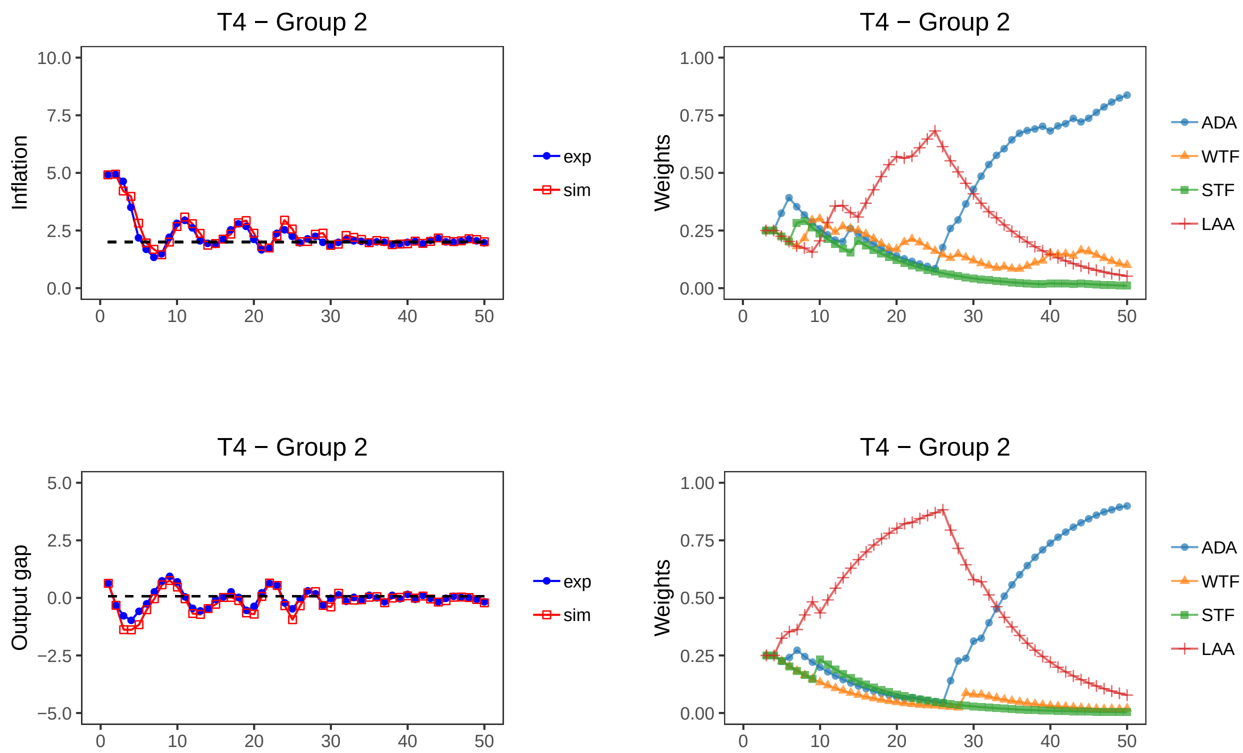
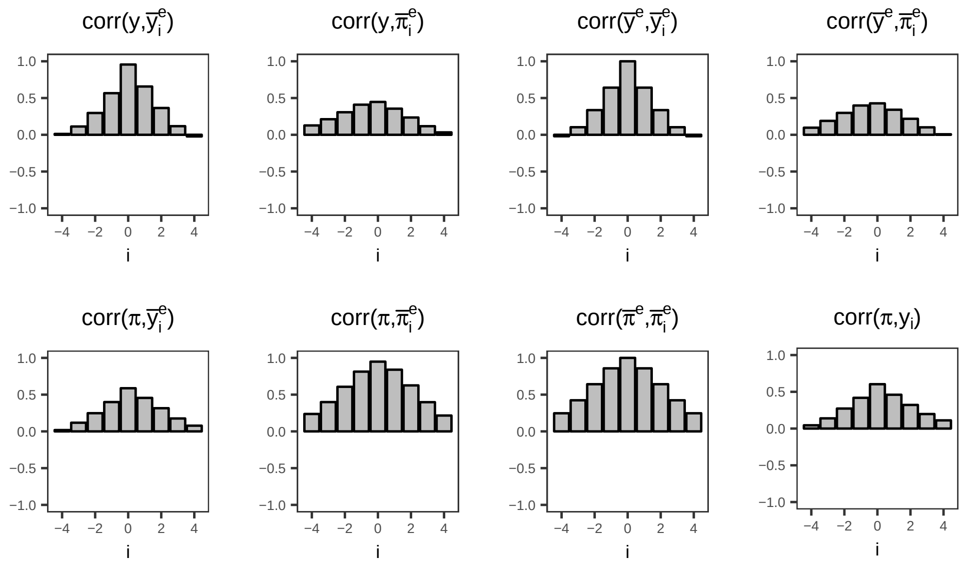
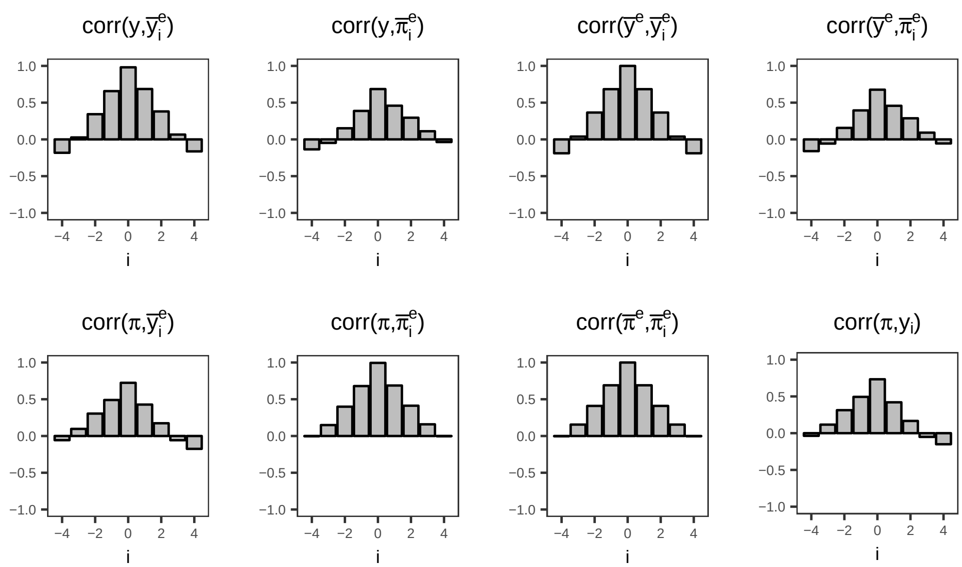


Figure 4: Realized and simulated inflation and output gap (left panels) with corresponding weights of 4 heuristics for $T4$ group 2. In the left panels, blue circles refer to experimental data while red squares refer to simulated data. In the right panels, ADA, WTR, STR and LAA refer respectively to the adaptive rule, the weak trend-following rule, the strong trend-following rule and the anchoring and adjustment rule.

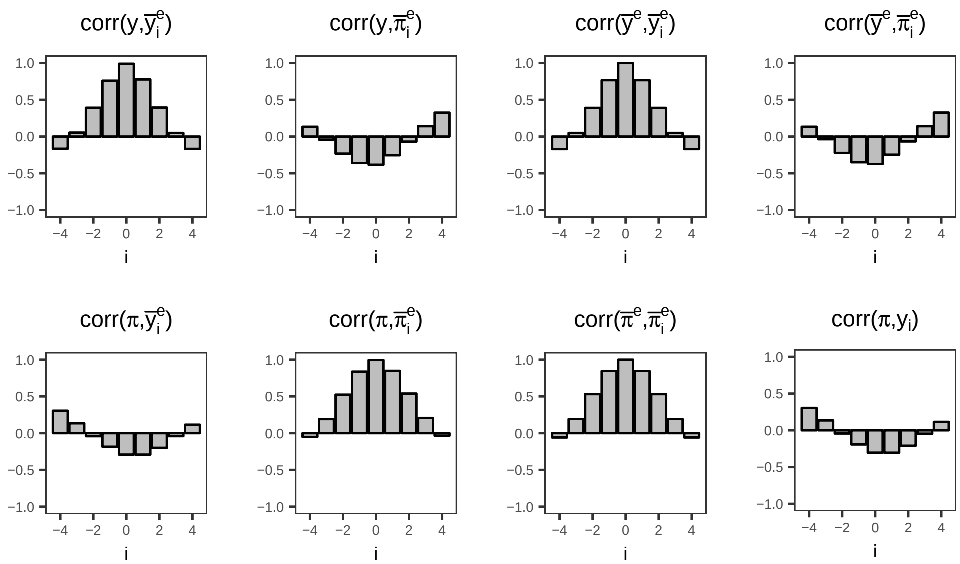


(a) Treatment $T1$

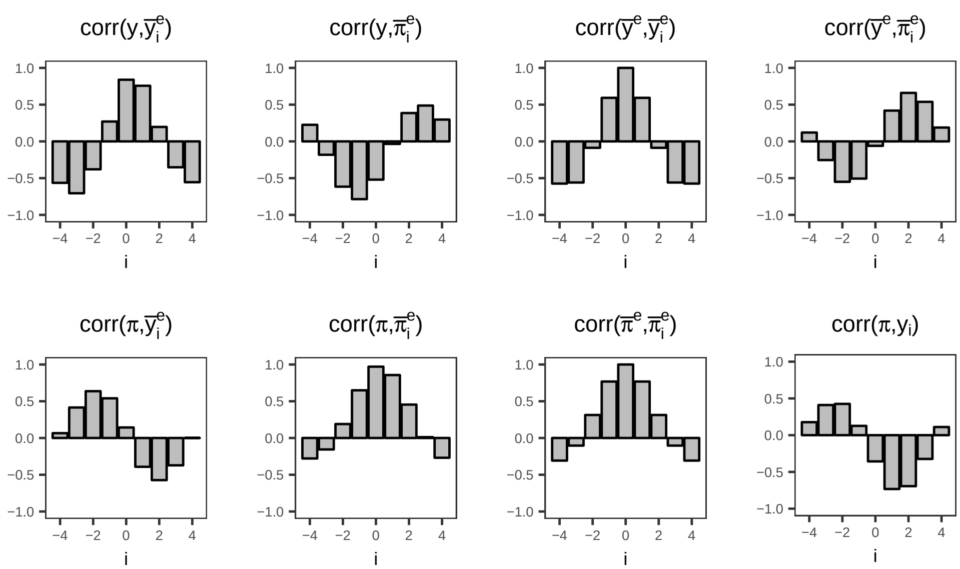


(b) Treatment $T2$

Figure 5: Correlations in experimental data – Purely positive feedbacks.



(a) Treatment $T3$



(b) Treatment $T4$

Figure 6: Correlations in experimental data – Mix positive/negative feedbacks.

Supplementary material (for online publication)

Appendix A. Instructions for participants

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 and output gap predictions of the statistical research bureaus in the economy (a group of six participants in this experiment) 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 15%.

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

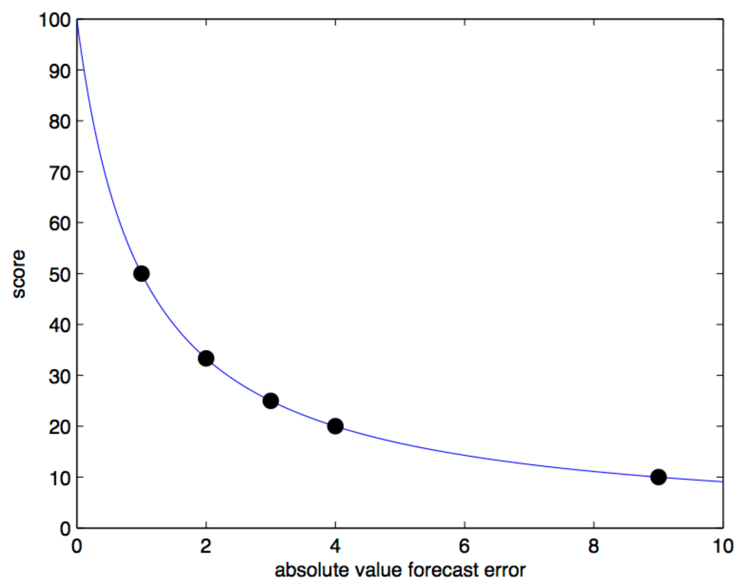
Your payment will depend on the accuracy of your predictions. You will be paid either for predicting inflation or for predicting the output gap. The accuracy of your predictions is measured by the absolute distance between your prediction and the actual values (this distance is the prediction error). 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

2

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.



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.**

Next to each graph, you will find an input box for your predictions.

On top of the **inflation** graph you are asked to enter your prediction for **inflation**.

At the bottom of the **output gap** graph you are asked to enter your prediction for the **output gap**.

In the bottom left part of the screen you will find a **Submit** button, to submit 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.

Appendix B. Computer interface

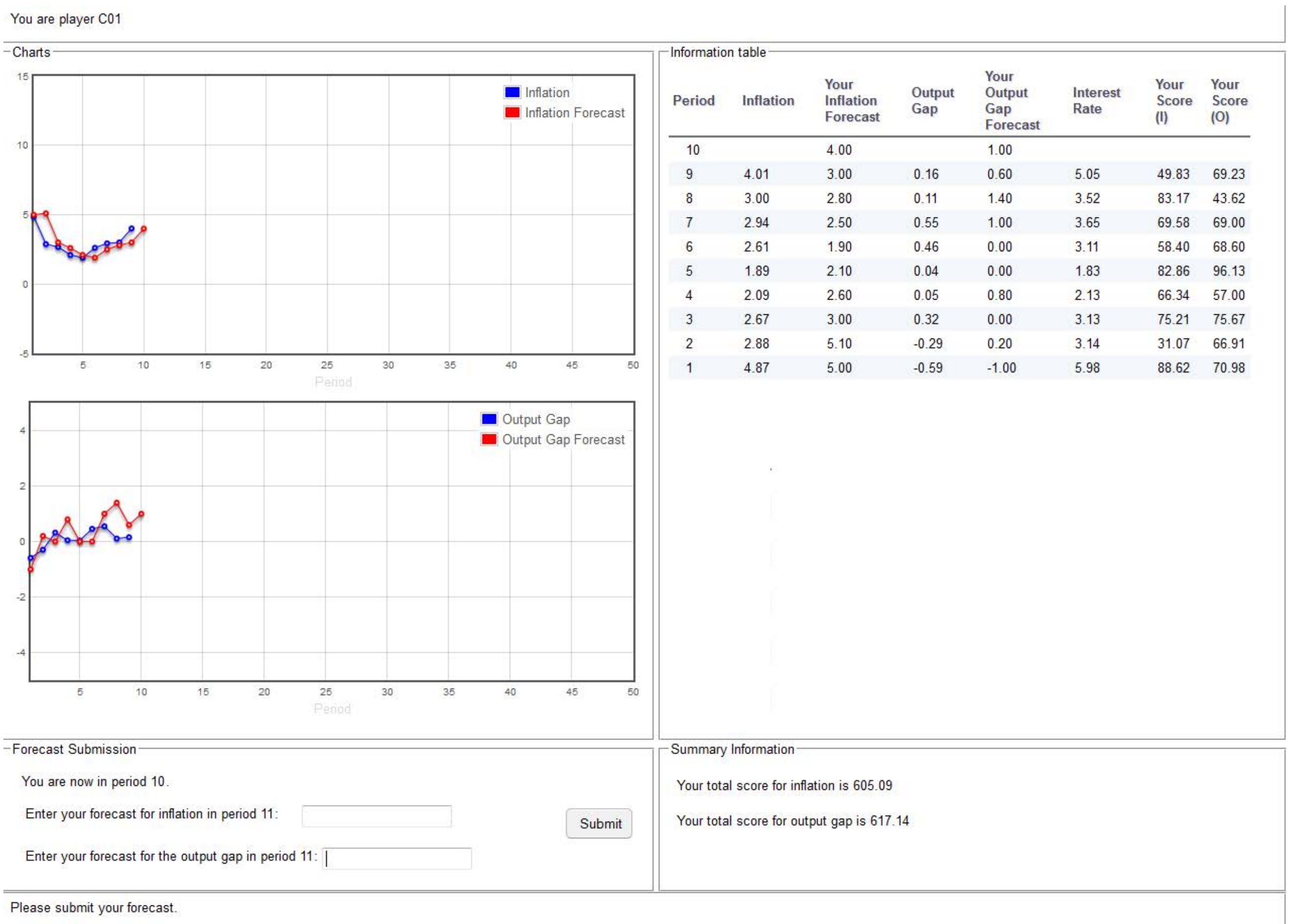


Figure B.7: Screenshot of computer interface.

Appendix C. Summary of all experimental data by group

Figs. C.8–C.15 show the realizations and forecasts of inflation and output gap. Each graph corresponds to one group of six people. The solid black line shows the realization of inflation (left panels) and the output gap (right panels), while the different markers show the forecasts of the six individuals in the group. For some experimental economies, for which dynamics were not very visible in the plot range $(-100, +100)$, we report a zoom over a smaller interval in the inset graphs.

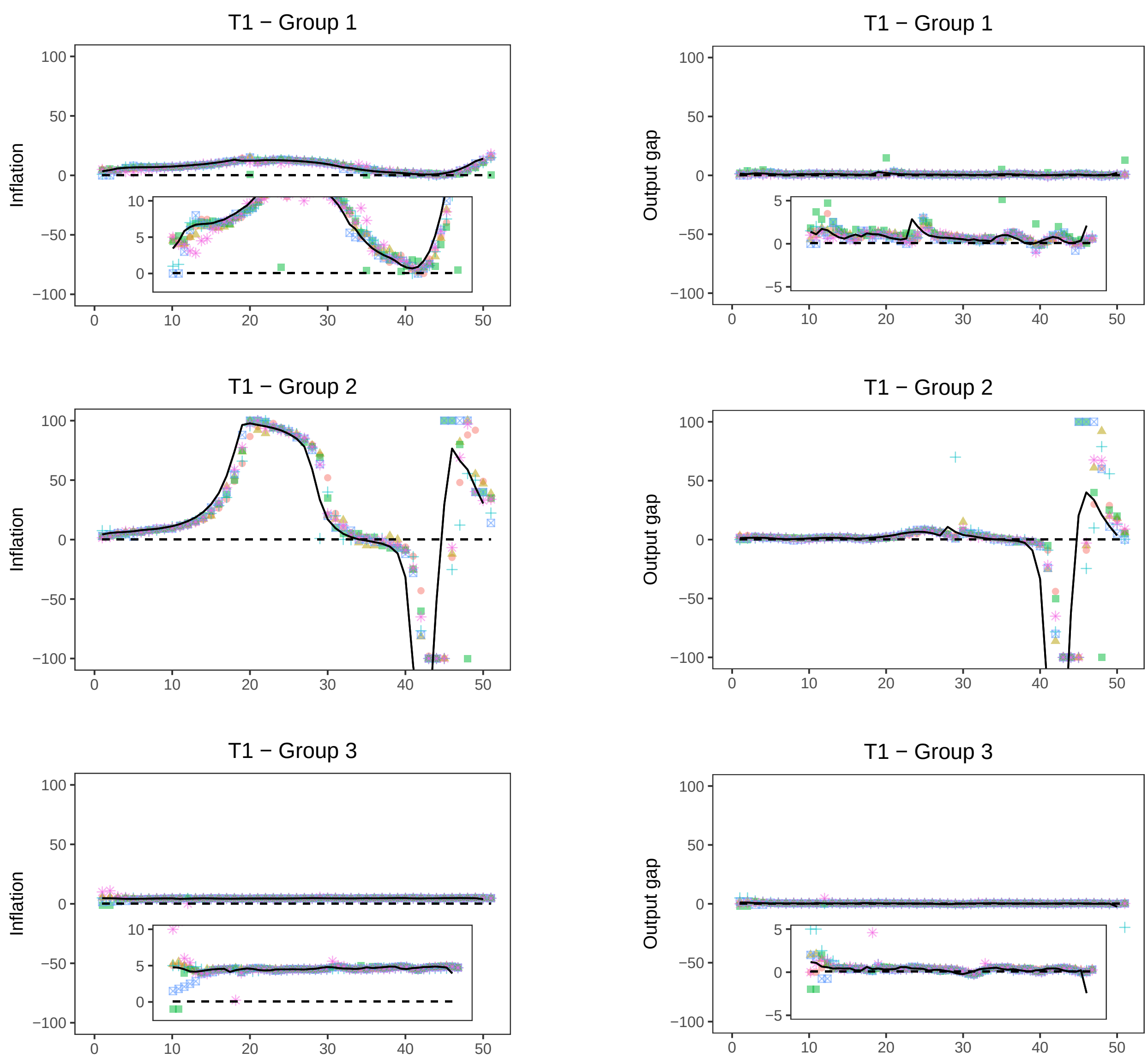


Figure C.8: Realizations and forecasts of inflation (left panels) and the output gap (right panels) for $T1$ (groups 1–3).

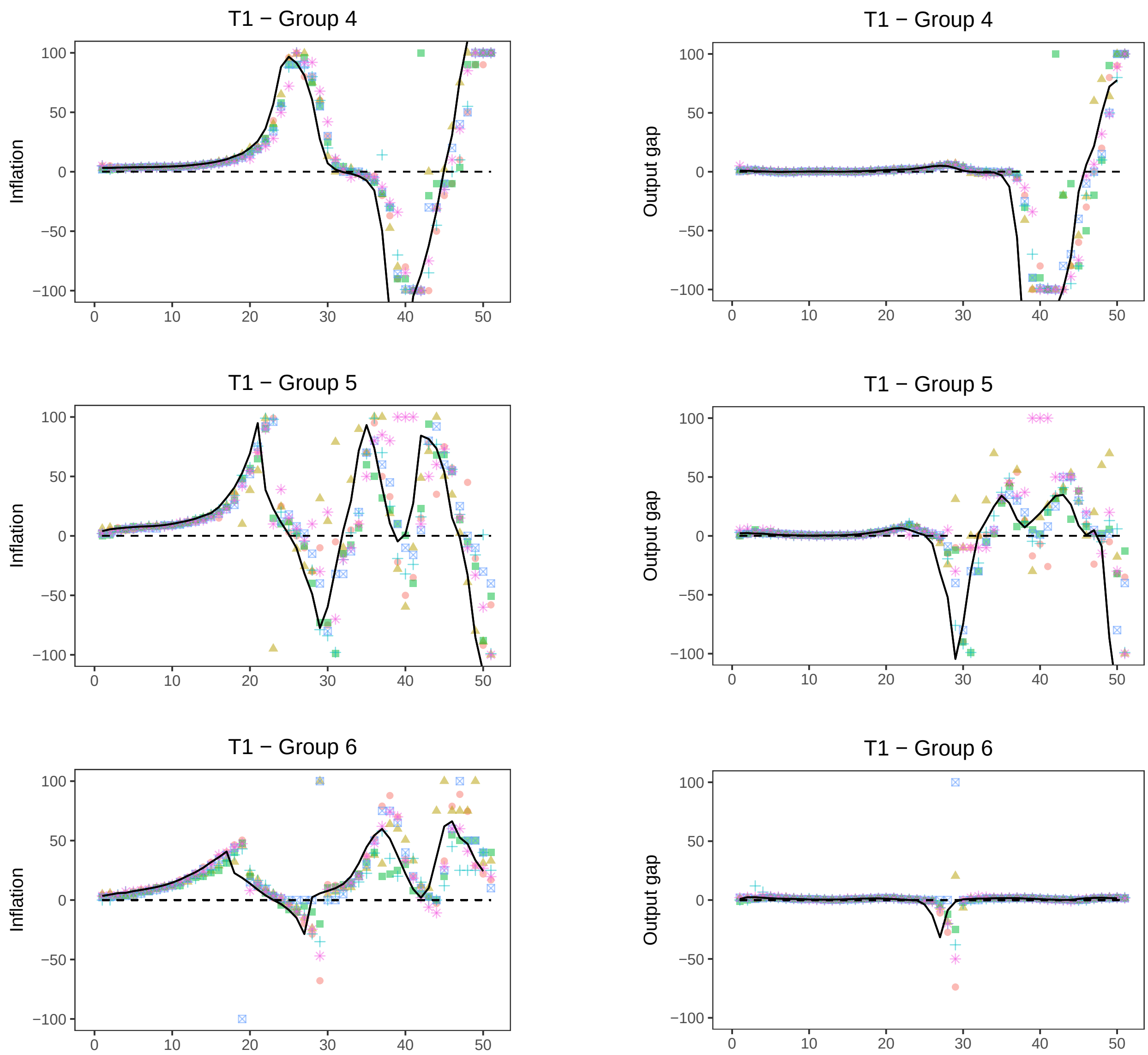


Figure C.9: Realizations and forecasts of inflation (left panels) and the output gap (right panels) for $T1$ (groups 4–6).

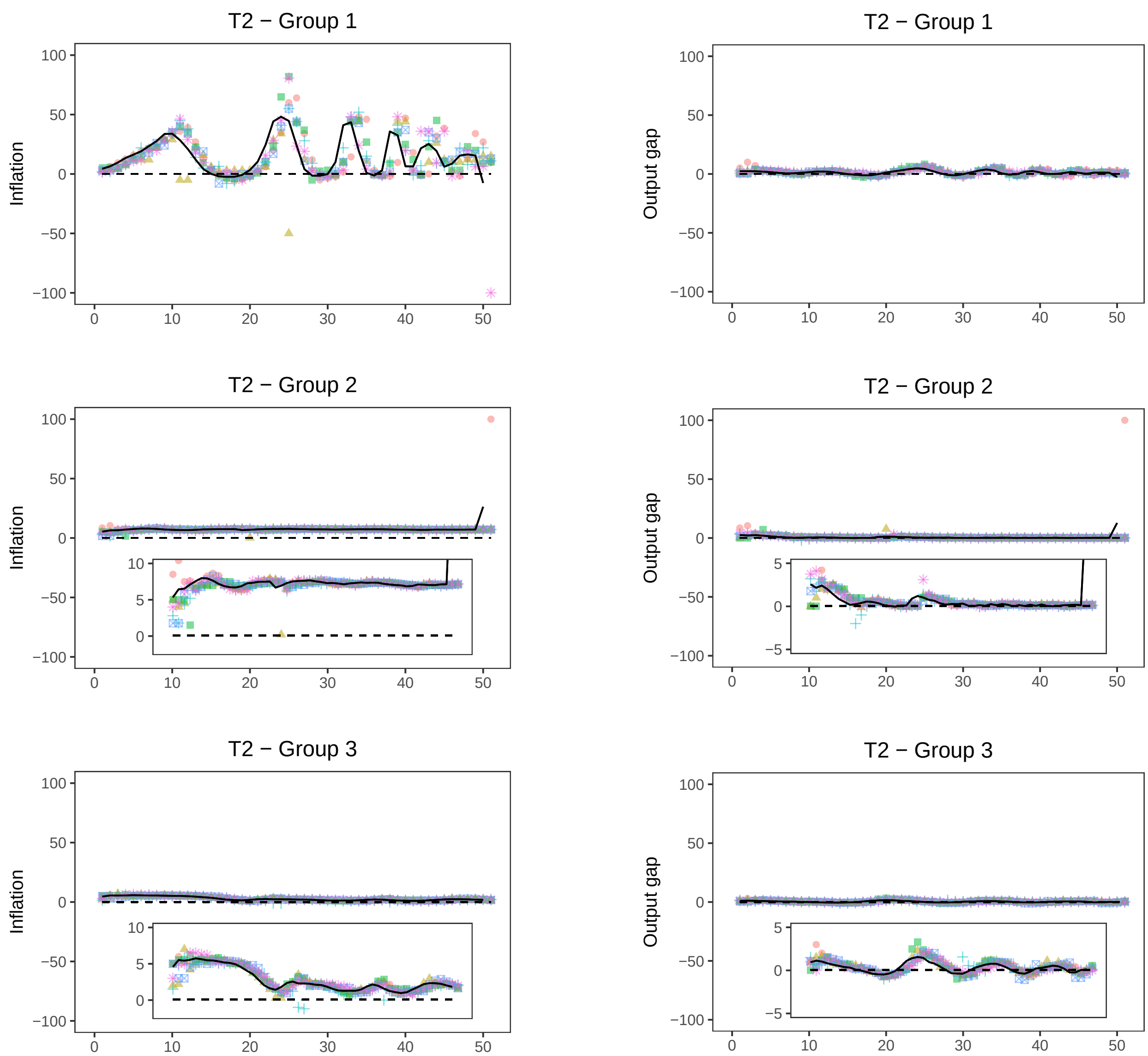


Figure C.10: Realizations and forecasts of inflation (left panels) and the output gap (right panels) for $T2$ (groups 1–3).

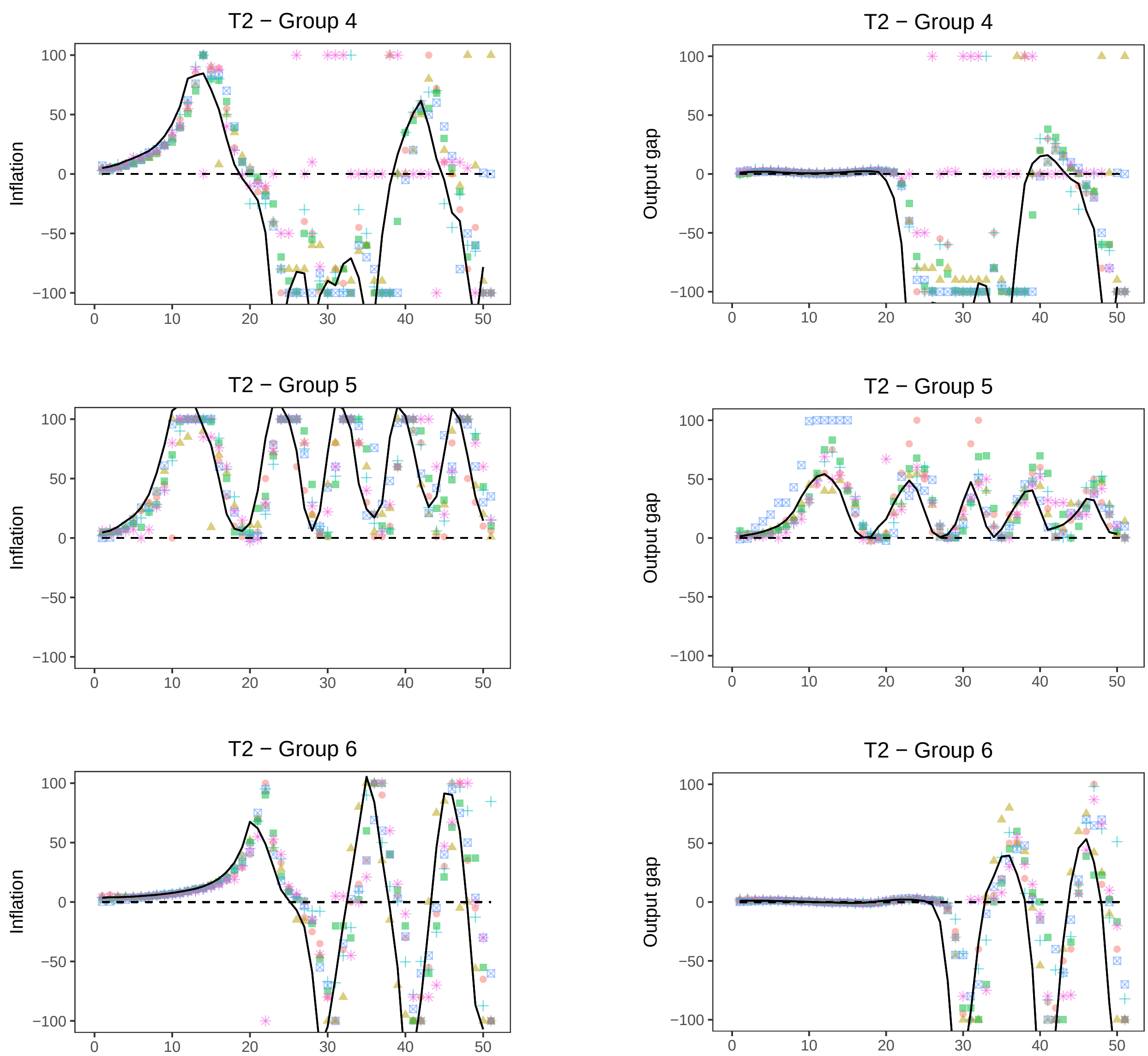


Figure C.11: Realizations and forecasts of inflation (left panels) and the output gap (right panels) for $T2$ (groups 4–6).

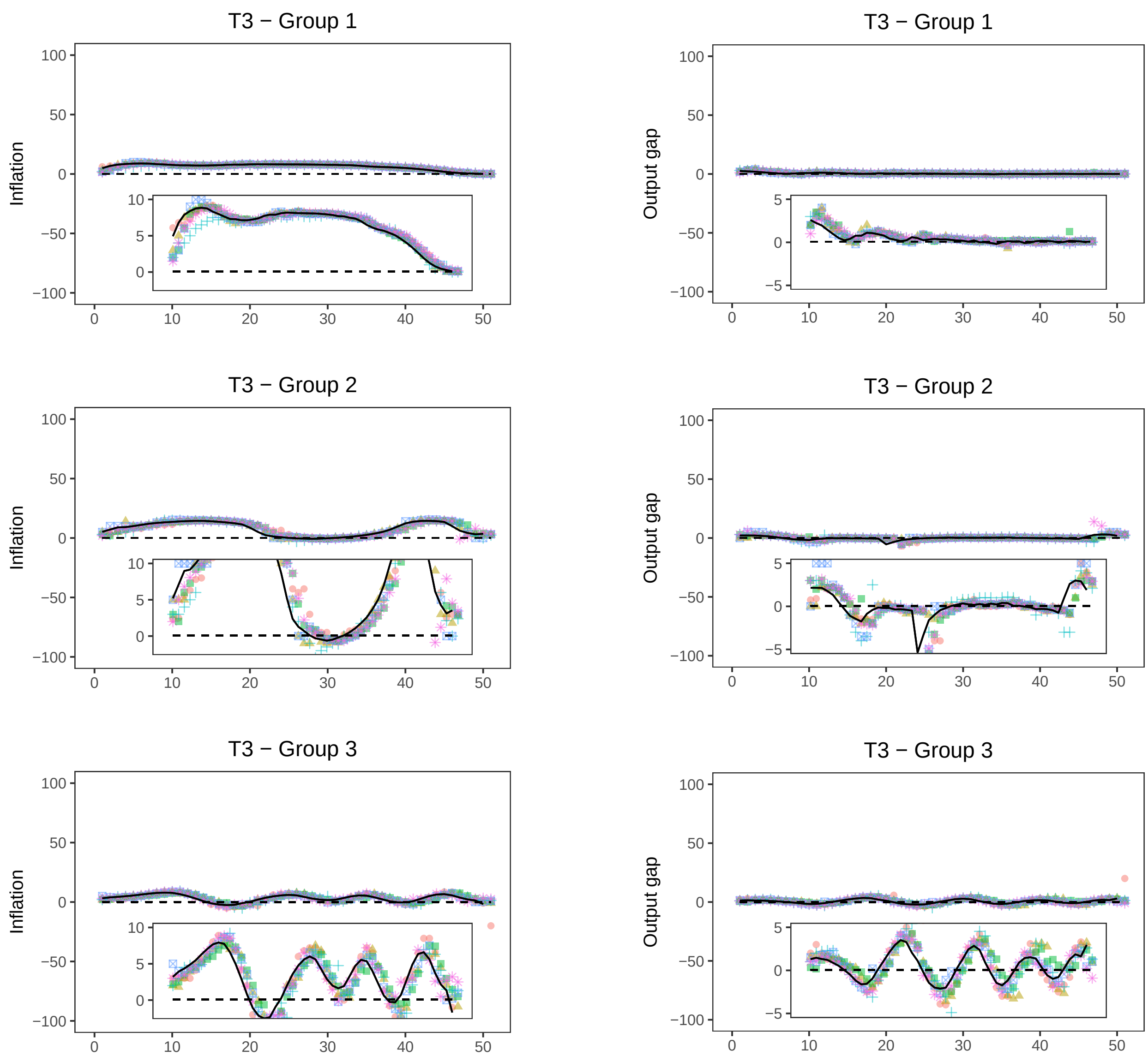


Figure C.12: Realizations and forecasts of inflation (left panels) and the output gap (right panels) for $T3$ (groups 1–3).

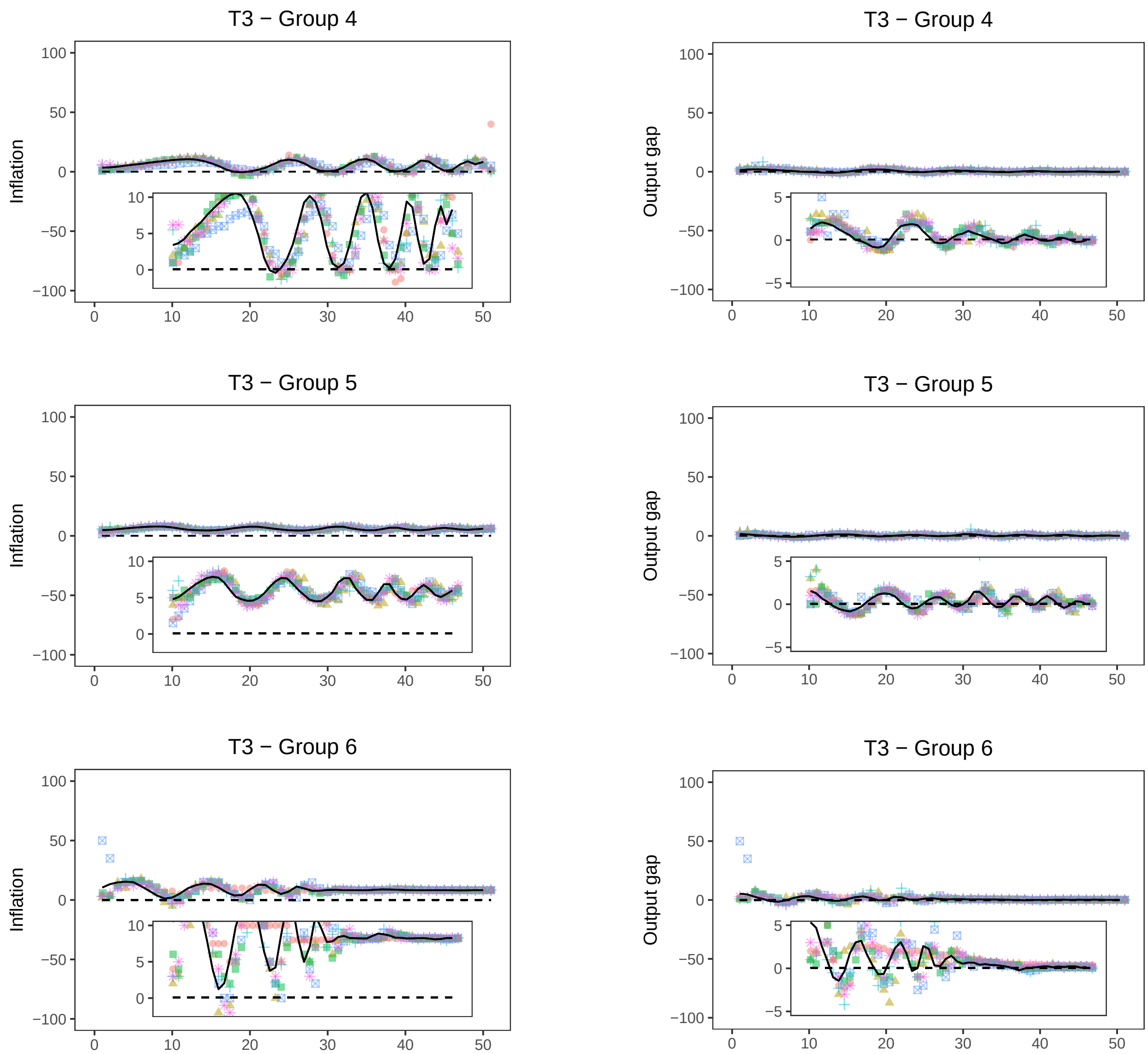


Figure C.13: Realizations and forecasts of inflation (left panels) and the output gap (right panels) for $T3$ (groups 4–6).

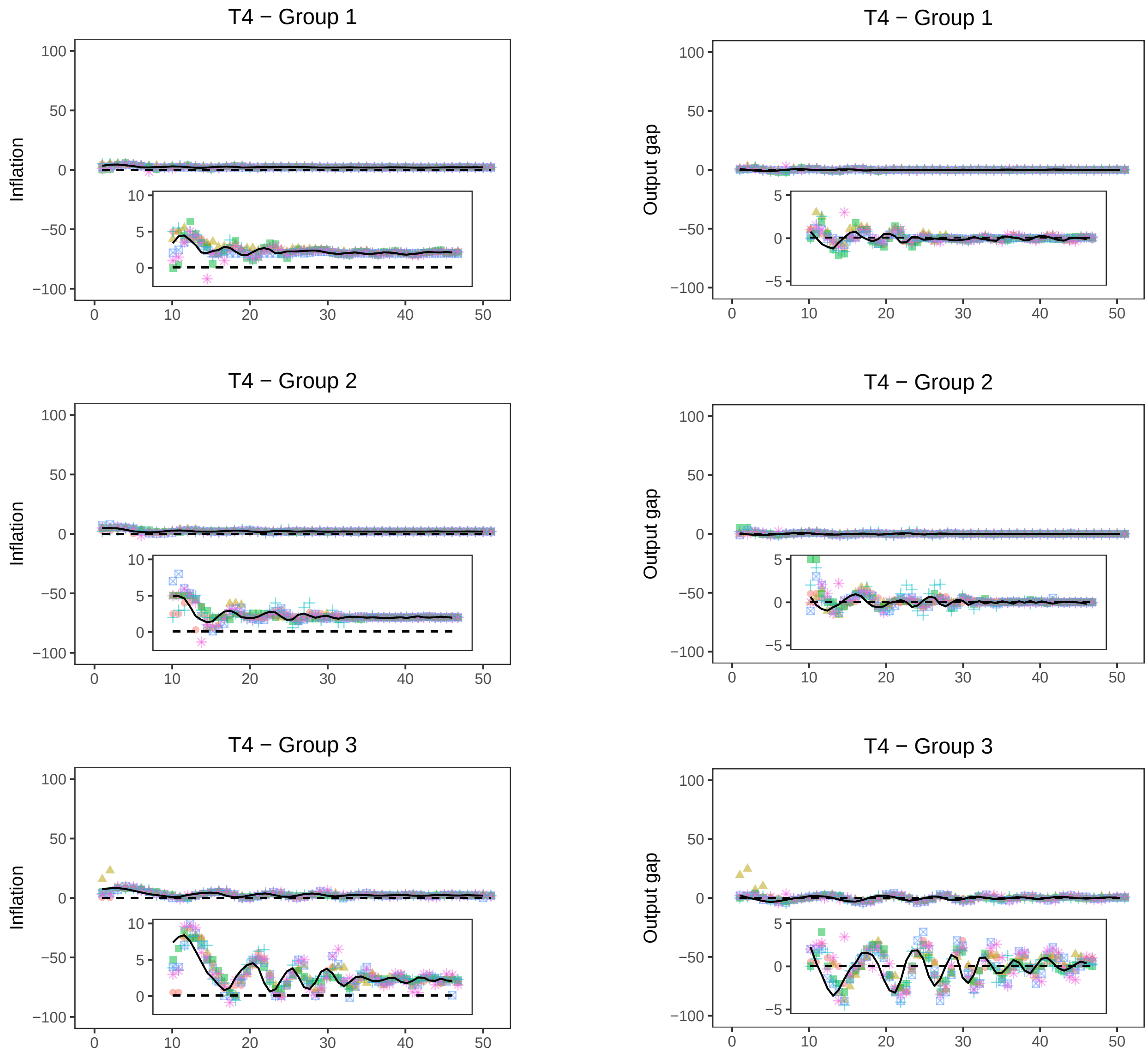


Figure C.14: Realizations and forecasts of inflation (left panels) and the output gap (right panels) for $T4$ (groups 1–3).

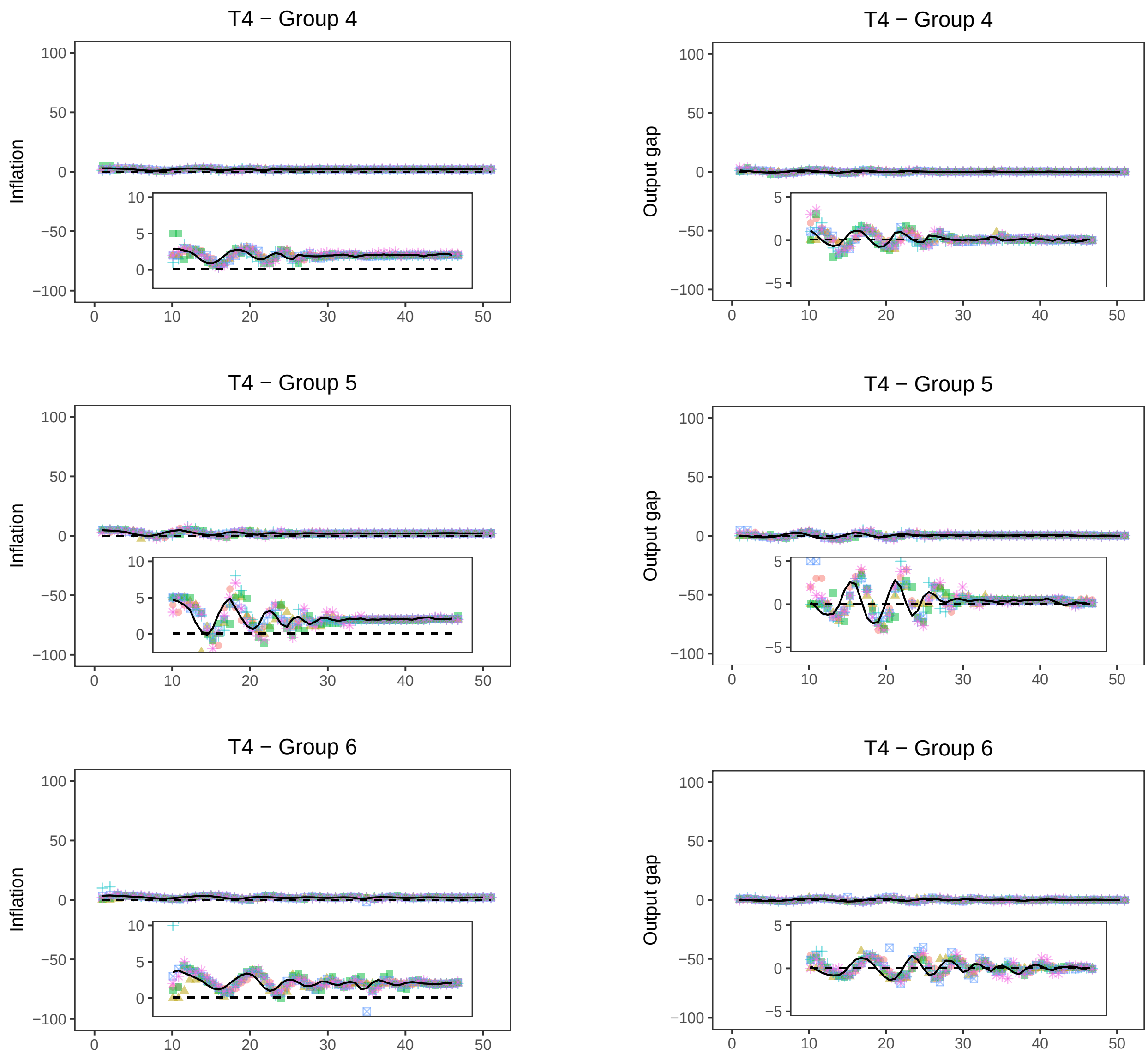


Figure C.15: Realizations and forecasts of inflation (left panels) and the output gap (right panels) for $T4$ (groups 4–6).

Figs. C.8–C.15 show that subjects tend to coordinate on a common prediction strategy, although participants in different groups may coordinate on different strategies. In order to quantify coordination on a common prediction strategy we consider, for each group, the average individual quadratic forecast error

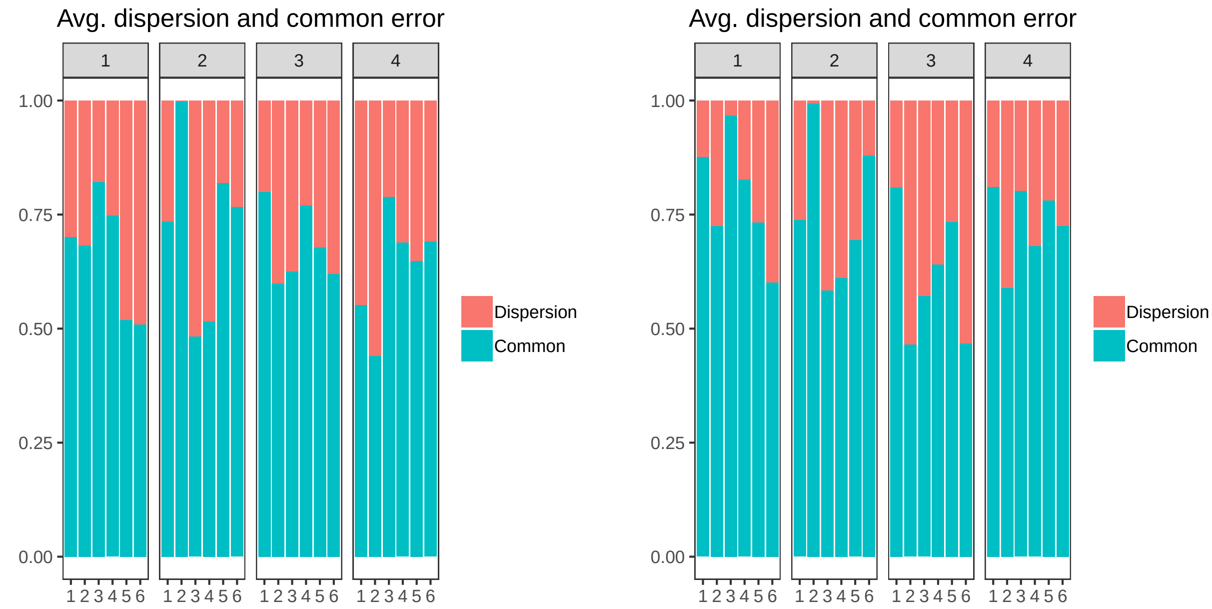
$$\frac{1}{6 \times 36} \sum_{i=1}^{6} \sum_{t=15}^{50} (x_{i,t}^e - x_t)^2 \ ,$$

defined as the individual quadratic forecast error averaged over time and over participants within a group. Note that we consider observations from period 15 on to allow for an initial learning phase. In the context of the NK model, x refers to either inflation or the output gap. Defining $\bar{x}_{i,t}^e = \sum_{i=1}^{6} x_{i,t}^e$ as the average prediction in a group, we can decompose the average individual quadratic forecast error as follows

$$\frac{1}{6 \times 36} \sum_{i=1}^{6} \sum_{t=15}^{50} (x_{i,t}^e - x_t)^2 \ = \frac{1}{6 \times 36} \sum_{i=1}^{6} \sum_{t=15}^{50} (x_{i,t}^e - \bar{x}_t^e)^2 + \frac{1}{36} \sum_{t=15}^{50} (\bar{x}_t^e - x_t)^2 \ . \qquad \text{(C.1)}$$

The first term on the RHS of Eq. (C.1) measures the dispersion among individual predictions as the quadratic distance between individual and average prediction within each group, averaged over time and participants. This term equals 0 when all participants in a group use exactly the same forecasting strategy. Therefore this term measures deviation from coordination on a common prediction strategy. The second term on the RHS of Eq. (C.1) measures instead the average distance between average forecast $\bar{x}_t^e$ and realization x_t. Fig. C.16 reports, for each of the 6 groups in the 4 treatments, the decomposition of the average quadratic forecast error into average dispersion and average common error.[25] From inspection of Fig. C.16 it is clear that only a relatively small part of the average quadratic forecasting error can be explained by the dispersion in expectations. In fact, on average respectively 68% and 72% of the average quadratic forecast error in inflation and output gap can be attributed to the average common error. Overall, the decomposition of the average quadratic forecast

[25]In order to avoid the big impact that outliers have on the measure of dispersion in individual forecasts, e.g. when bounds on individual predictions were reached or subjects tried to strategically reverse explosive trends, we remove them using linear interpolation of neighboring, non-outlier values. Outliers are defined as observations more than three MAD from the local median defined over a window of 4 observations.



(a) Inflation

(b) Output gap

Figure C.16: Decomposition of average quadratic forecast error of individual prediction strategies into average dispersion error and average common error for each of the 6 groups in the 4 treatments.

error suggests that there is coordination on a common prediction strategy, although some heterogeneity in individual forecasts persists.

Appendix D. Switching behavior

Evidence of switching behavior can be found by inspecting the time series of individual forecasts. Fig. D.17 reports some graphical evidence of individual switching behavior. For every period t we plot realized inflation or output gap in that period, together with the prediction submitted by subjects in period $t + 1$. In this way we can graphically infer how individual predictions use past available observations of the variable being forecasted. For example, if the time series coincide, the participant is submitting predictions identical to the last observation.

In Fig. D.17(a) (treatment $T4$, group 2), subject 12 extrapolates the direction of change in inflation in the early stage of the experiment. Starting from about period 20 the participant switches to a much weaker form of trend extrapolation, to later on adopt an adaptive forecasting strategy in which individual forecasts are somewhere in between the last available observation and the previous prediction.

In Fig. D.17(b) (treatment $T4$, group 2), we observe a somewhat similar forecasting behavior as subject 11 strongly extrapolates past changes in the output gap in the first half of the experiment. In the second half the participant switches to an adaptive forecasting heuristic.

In Fig. D.17(c) (treatment $T3$, group 6), subject 7 switches between various constant predictors for inflation in the first 20 periods of the experimental session. Later on the participant converges to a predictor of about 8%, which represents an almost self-fulfilling equilibrium for the experimental economy.

In Fig. D.17(d) (treatment $T3$, group 6), subject 12 starts out with a trend extrapolating strategy and later on switches to a "naive" forecasting rule that basically uses the last available observation to predict future output gap.

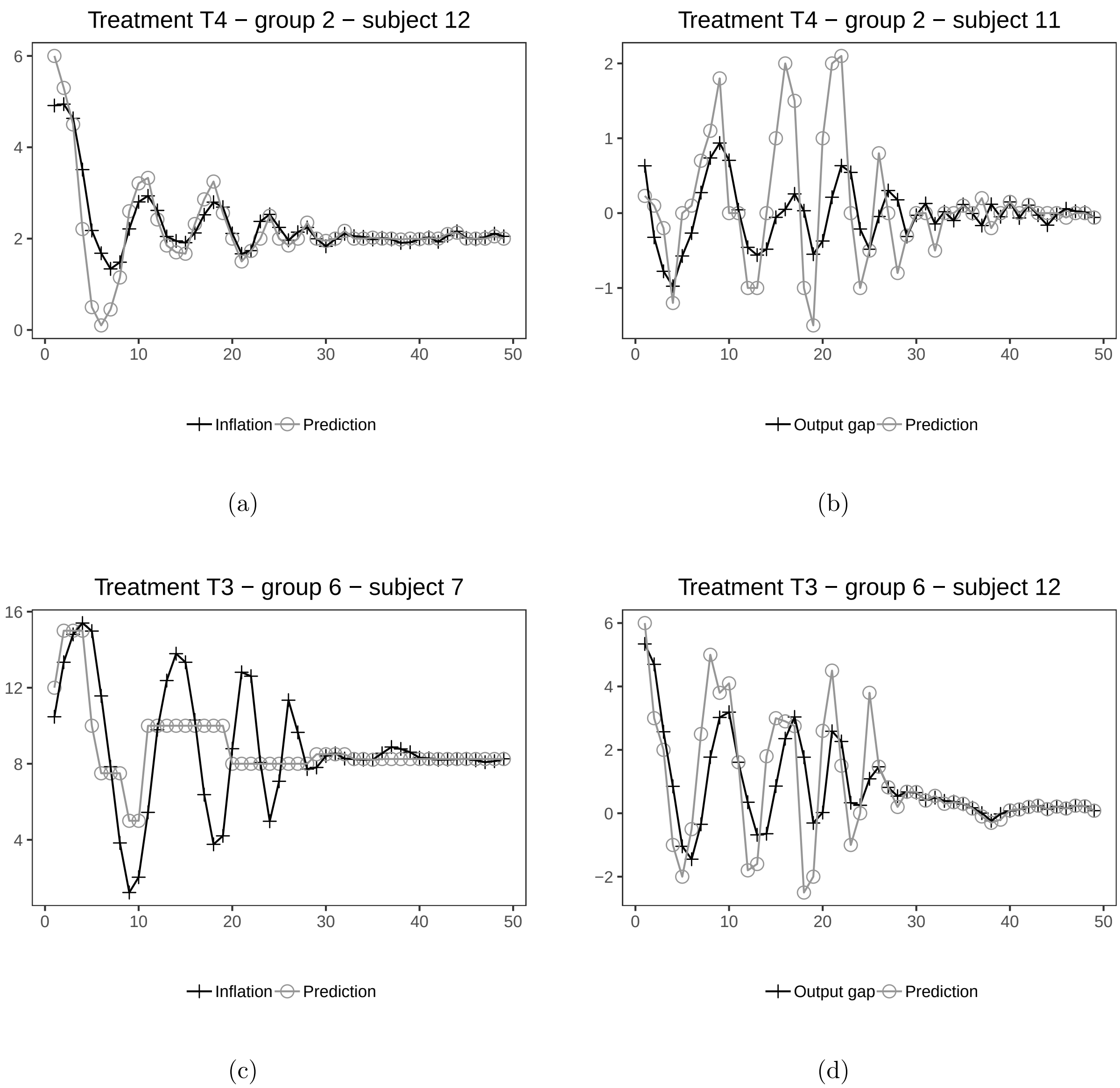


Figure D.17: Individual learning as switching between heuristics. For every period t, subject i's prediction $x^e_{i,t+2}$ and the last available observation of the variable x_t being forecasted (with x being either inflation or the output gap) are reported.

Appendix E. Estimation of forecasting rules

In what follows we only consider experimental economies in which expectations did not reach the artificial bounds on admissible forecasts. Accordingly, we exclude from the sample groups 2, 4, 5 and 6 in $T1$ and groups 4, 5, and 6 in $T2$. In all analyses performed below, we consider a significance level of 0.05. For each of the 102 participants in the considered subsample we estimated linear prediction rules of the form

$$\pi^e_{j,t+1} = c + \sum_{i=0}^{2} \alpha^e_i \pi^e_{j,t-i} + \sum_{i=1}^{3} \alpha^\pi_i \pi_{t-i} + \sum_{i=1}^{3} \alpha^y_i y_{t-i} + \xi_t \tag{E.1}$$

$$y^e_{j,t+1} = c + \sum_{i=0}^{2} \alpha^e_i y^e_{j,t-i} + \sum_{i=1}^{3} \alpha^y_i y_{t-i} + \sum_{i=1}^{3} \alpha^\pi_i \pi_{t-i} + \epsilon_t \ , \tag{E.2}$$

where $\pi^e_{j,t+1}$ and $y^e_{j,t+1}$ refer to inflation or output gap forecast of participant j for period $t+1$ (submitted in period t). We allow for an initial learning phase, in which subjects may have not yet converged to a prediction rule and still be experimenting with different strategies, by considering observations starting from period 15.[26] Overall, for about 65% of the estimated rules we do not detect any first-order autocorrelation in the residuals according to a Breusch-Godfrey test. Moreover, in about 75% of the cases an $F-$test indicates that we can restrict rules (E.1)–(E.2) to simpler rules in which predictions depend only on past forecasts and past observations of the forecasting objective. Averaging over participants of all treatments, the number of significant regressors in the estimated prediction rules is about 2. The most popular significant regressor is the last available observation of the forecasting objective (π_{t-1} or y_{t-1}), followed by the second last available observation π_{t-2} for Eq. (E.1) and by the most recent own prediction y^e_t for Eq. (E.2). Looking at the estimated coefficients, a remarkable property is that 100% of the significant coefficients associated to the last observed forecasting objective and about 90% of the significant coefficients associated to the most recent own prediction are positive. In contrast, about 92% of the significant coefficients associated to the second last observed forecasting objective are negative.

[26] We remove outliers in individual forecasts using linear interpolation of neighboring, non-outlier values. Outliers are defined as observations more than three MAD from the local median defined over a window of 4 observations.

Overall, the estimation results indicate that most participants use a linear prediction rule, at least after an initial learning phase. What is more, the fact that the two latest observations of the forecasting objective and the latest own prediction are generally the most used prediction rule components, implies that these variables are of particular importance in the prediction rule specification. The relatively low average number of significant regressors in Eqs. (E.1)–(E.2) means that the other variables are used very little as input to form predictions. It is therefore worthwhile to restrict specifications (E.1)–(E.2) by leaving out these infrequently used regressors. The fact that the estimated non-zero coefficients for the most recent values of the forecasting objective and the own prediction are almost all positive, while the non-zero coefficients of the other variables tend to be negative, similarly suggests that the specifications (E.1)–(E.2) are too flexible and could be restricted without losing much explanatory power. Restricting (E.1)–(E.2) along the lines of these regularities could increase the efficiency of the estimates, as well as make the estimated rules easier to interpret from a behavioral viewpoint.

In particular, we perform an $F-$test to check whether we could restrict the general forecasting rules in Eqs. (E.1)–(E.2) to simpler prediction rules of the form

$$\pi^e_{j,t+1} = \alpha_1 \pi_{t-1} + \alpha_2 \pi^e_{j,t} + (1 - \alpha_1 - \alpha_2) \frac{1}{35} \sum_{t=15}^{50} \pi_t + \alpha_3 (\pi_{t-1} - \pi_{t-2}) + \xi_t \qquad \text{(E.3)}$$

$$y^e_{j,t+1} = \alpha_1 y_{t-1} + \alpha_2 y^e_{j,t} + (1 - \alpha_1 - \alpha_2) \frac{1}{35} \sum_{t=15}^{50} y_t + \alpha_3 (y_{t-1} - y_{t-2}) + \epsilon_t \,. \qquad \text{(E.4)}$$

Forecasting rules (E.3)–(E.4) are referred to as *First-Order Heuristics* (FOH) and can be interpreted as anchoring-and-adjustment heuristics *à la* Tversky and Kahneman. The first three terms in (E.3) and (E.4) are a weighted average of the latest realization of the forecasting objective, the latest own prediction and the forecasting objective's sample mean (excluding a learning phase).[27] This weighted average is the (time varying) "anchor" of the prediction, which is a zeroth-order extrapolation from the available data at period t. The

[27] In the estimation of (E.3) and (E.4) we include the sample mean of inflation and the output gap, which is of course not available to the subjects at the moment of the prediction, but acts as a proxy of the equilibrium level. In the HSM of Section 4.1, the LAA rule uses sample average up to $t-1$, which is observable to subjects when the forecast is made and generally converges quickly to the full sample mean.

fourth term in (E.3) and (E.4) is a simple first-order extrapolation from the two most recent realizations of the forecasting objective; this term is the "adjustment" or trend extrapolation part of the heuristic. An advantage of FOH is that it simplifies to well-known forecasting rules for different boundary values of the parameter space. For example, Eqs. (E.3)–(E.4) reduce to Naive Expectations if $\alpha_1 = 1$, $\alpha_2 = \alpha_3 = 0$; they reduce to Adaptive Expectations if $\alpha_1 + \alpha_2 = 1$ (with $\alpha_1, \alpha_2 \in (0, 1)$) and $\alpha_3 = 0$ (ADA rule considered in Section 4.1); they reduce to the simplest Trend-Following rule if $\alpha_1 = 1$, $\alpha_2 = 0$ and $\alpha_3 > 0$ (WTR and STR rules considered in Section 4.1). When $0 < \alpha_1 < 1$, $\alpha_2 = 0$ and $\alpha_3 = 1$, with the sample average computed using observations up to period $t - 1$, we obtain an Anchoring and Adjustment rule with a time-varying anchor (LAA rule considered in Section 4.1). Overall, about 66% of the general forecasting rules (E.1)–(E.2) could be restricted to FOH rules (E.3)–(E.4) according to an $F-$test. In about 54% of the cases we do not detect any first-order autocorrelation in the residuals according to a Breusch-Godfrey test. Moreover, about 53% of the estimated rules could be exactly restricted to one of the types considered in Section 4.1, while the others present different anchor–adjustment combinations within the classes defined in Eqs. (E.3)–(E.4). Fig. E.18 reports estimates of the FOH coefficients in Eqs. (E.3)–(E.4).

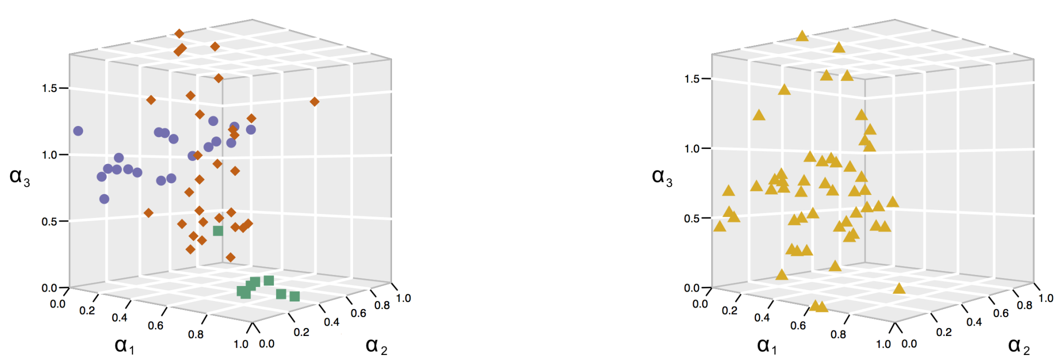


Figure E.18: Left panel: estimated coefficients of rules classified as Adaptive (squares), Trend-Following (diamonds) and Anchoring and Adjustment (circles). Right panel: estimated coefficients of rules with different anchor–adjustment combinations.

Appendix F. Homogeneous expectations models

In this section we analyze the stability properties of the NK model in Eq. (4) under homogeneous expectations, i.e. when all participants in the economy use the same forecasting heuristic. In particular, we study the deterministic skeleton of model (4), i.e. setting the noise term ϵ_t to zero, under the homogeneous expectations presented in Table 2 in different policy regimes.

Adaptive heuristics

Under adaptive expectations for both inflation and the output gap we can write the vector of expected future aggregate variables $z^e = (y^e, \pi^e)'$ as

$$z^e_{t+1} = \chi z_{t-1} + (1-\chi) z^e_t \ , \tag{F.1}$$

where scalar $0 < \chi < 1$ denotes the relative weight of past observations. Rewriting the NK model in Eq. (4) as

$$z^e_{t+1} = -\mathrm{M}^{-1}\mathrm{A} + \mathrm{M}^{-1} z_t \ . \tag{F.2}$$

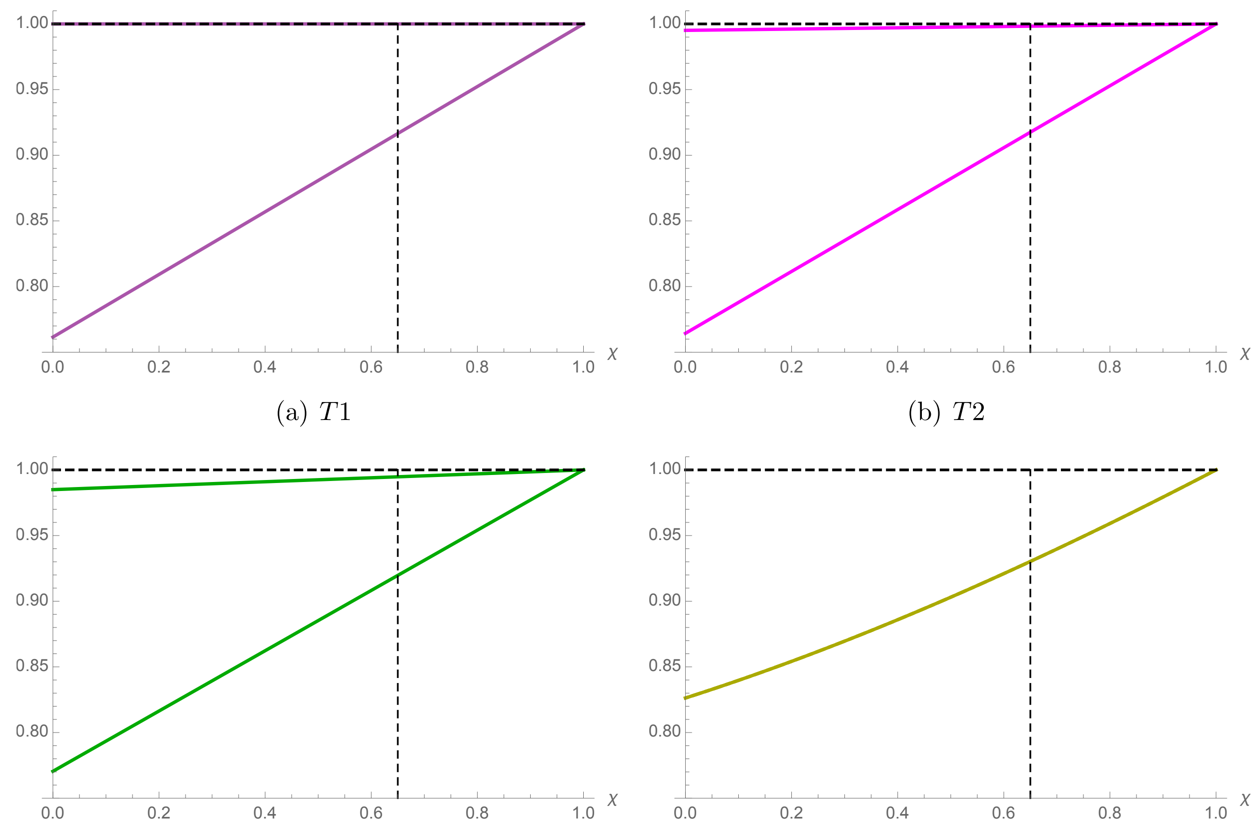
Substituting Eq. (F.2) lagged one period in Eq. (F.1) we can write z^e_{t+1} as function of z_{t-1}

$$z^e_{t+1} = -(1-\chi)\,\mathrm{M}^{-1}\mathrm{A} + (\chi\, I + (1-\chi)\,\mathrm{M}^{-1}) z_{t-1} \ , \tag{F.3}$$

where I denotes the identity matrix. Substituting Eq. (F.3) in the NK model (4) we obtain

$$z_t = \chi\,\mathrm{A} + (\chi\,\mathrm{M} + (1-\chi)\, I) z_{t-1} \ . \tag{F.4}$$

The dynamic properties of the NK model under homogeneous adaptive expectations are described by Eq. (F.4). Simple calculations show that the unique steady state of system (F.4) is the FS-RE equilibrium $\bar{z} = (I - \mathrm{M})^{-1}\mathrm{A}$, provided that matrix $(I - \mathrm{M})$ is invertible, i.e. $\phi_\pi > 1$. Stability of the FS-RE equilibrium under adaptive expectations depends on the eigenvalues of matrix $\chi\,\mathrm{M} + (1-\chi)\, I$. Fig. F.19 displays the absolute value of the eigenvalues of matrix $\chi\,\mathrm{M} + (1-\chi)\, I$ as function of parameter χ under policy regimes implemented in different treatments. In treatment $T1$ one eigenvalue is always on the unit circle so that



(a) $T1$ (b) $T2$

(c) $T3$ (d) $T4$

Figure F.19: Absolute value of eigenvalues of matrix $\chi\,\mathrm{M} + (1-\chi)\,I$ as function of χ. Dashed vertical lines refer to ADA rule ($\chi = 0.65$).

the steady state of Eq. (F.4) is indeterminate and there is a continuum of stable steady states. In treatments $T2$, $T3$ and $T4$ the FS-RE steady state is stable for all values of χ. Convergence under homogeneous adaptive expectations is monotonic in $T2$ and $T3$ due to real eigenvalues and oscillatory in $T4$ due to complex eigenvalues.

Trend-following heuristics

Under trend-following heuristics for both inflation and the output gap we can write the vector of expected future aggregate variables $z^e = (y^e, \pi^e)'$ as

$$z^e_{t+1} = z_{t-1} + \xi(z_{t-1} - z_{t-2}) \ , \tag{F.5}$$

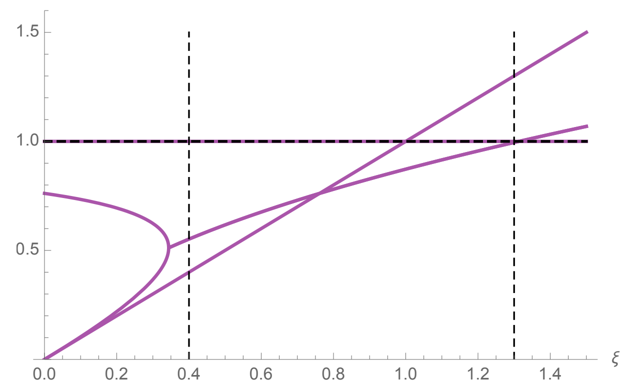
where scalar $\xi > 0$ denotes the degree of trend-extrapolation. Under expectations defined in Eq. (F.5) the NK model can be rewritten as

$$z_t = \mathrm{A} + \mathrm{M}\,(1+\xi) z_{t-1} - \xi\, z_{t-2} \ . \tag{F.6}$$

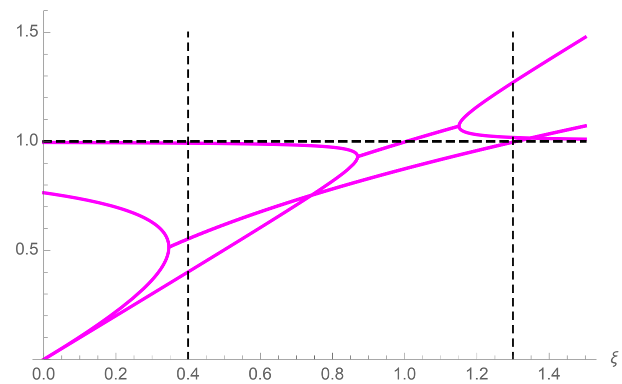
Defining $w_t = z_{t-1}$ we can rewrite Eq. (F.6) as a first-order system defined by

$$\begin{aligned} \begin{pmatrix} z_t \\ w_t \end{pmatrix} &= \begin{pmatrix} \mathrm{A} \\ 0 \end{pmatrix} + \begin{pmatrix} (1+\xi)\,\mathrm{M} & -\xi\,\mathrm{M} \\ I & 0 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ w_{t-1} \end{pmatrix} \\ s_t &= \mathrm{B} + \mathrm{N}\, s_{t-1} \ , \end{aligned} \tag{F.7}$$

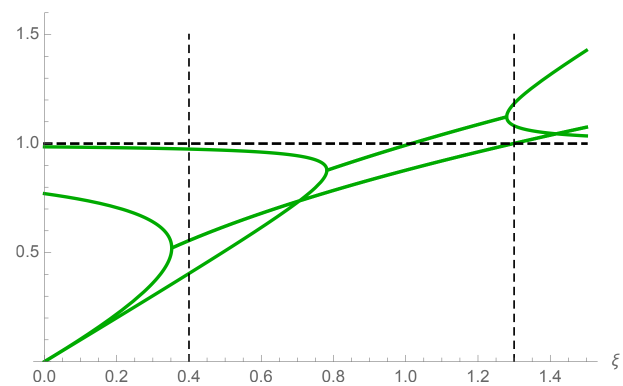
where $s = (z, w)'$. The dynamic properties of the NK model under homogeneous trend-following expectations are described by the 4-dimensional system in Eq. (F.7). Simple calculations show that the unique steady state of system (F.7) is the FS-RE equilibrium $\bar{z} = (I - \mathrm{M})^{-1}\mathrm{A}$, provided that matrix $(I - \mathrm{M})$ is invertible, i.e. $\phi_\pi > 1$. Stability of the FS-RE equilibrium under trend-following expectations depends on the eigenvalues of matrix N in Eq. (F.7). Fig. F.20 displays the absolute value of the eigenvalues of matrix N as function of parameter ξ under policy regimes of $T1$, $T2$, $T3$ and $T4$. Under the WTR in Table 2, i.e. $\xi = 0.4$, all eigenvalues are within the unit circle in $T2$, $T3$ and $T4$, meaning that the FS-RE steady state is stable (although convergence can be slow in $T2$ and $T3$ due to one eigenvalue close to one), while in $T1$ one eigenvalue is exactly on the unit circle, meaning



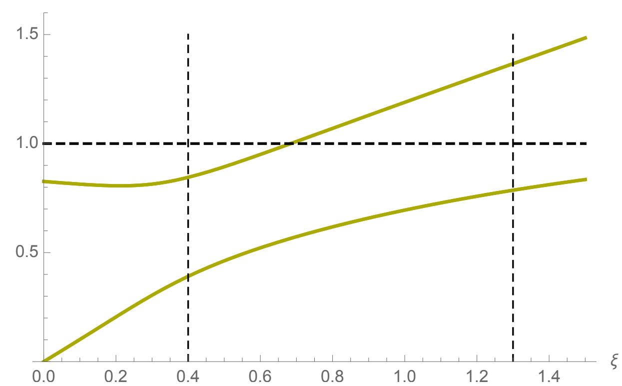
(a) $T1$



(b) $T2$



(c) $T3$



(d) $T4$

Figure F.20: Absolute value of eigenvalues of matrix N as function of ξ in different treatments. Dashed vertical lines refer to WTR ($\xi = 0.4$) and STR ($\xi = 1.3$).

that there is a continuum of stable equilibria. On the opposite, the system is unstable under the STR in Table 2, i.e. $\xi = 1.3$, in all treatments with dynamics exploding monotonically in treatments $T1$, $T2$ and $T3$ due to the presence of explosive real eigenvalues, and oscillating in $T4$ due to the complex explosive eigenvalues.

Anchoring and adjustment heuristics

The anchoring and adjustment heuristic considered in Table 2 for a generic variable x has a time-varying component $\bar{x}_{t-1}$ defined as

$$\bar{x}_{t-1} = \frac{1}{t-1} \sum_{i=1}^{t-1} x_i \ . \tag{F.8}$$

Therefore, under the anchoring and adjustment heuristics for both inflation and the output gap we can write the vector of expected future aggregate variables $z^e = (y^e, \pi^e)'$ as

$$z^e_{t+1} = \frac{1}{2}\bar{z}_{t-1} + \frac{3}{2}z_{t-1} - z_{t-2} \ . \tag{F.9}$$

Substituting Eq. (F.9) in the NK model we obtain

$$z_t = \mathrm{A} + \mathrm{M} \left(\frac{1}{2}\bar{z}_{t-1} + \frac{3}{2}z_{t-1} - z_{t-2} \right) \ . \tag{F.10}$$

Although it is trivial to show that that the FS-RE equilibrium $\bar{z} = (I - \mathrm{M})^{-1}\mathrm{A}$ is the unique steady state of system (F.10), provided that matrix $(I - \mathrm{M})$ is invertible, i.e. $\phi_\pi > 1$, it is non-trivial to study its stability properties due to explicit dependence on t. Therefore, we replace $\bar{z}_{t-1}$ with the equilibrium $\bar{z}$ and study whether small perturbations to the FS-RE equilibrium are amplified or re-absorbed. We thus consider the system

$$z_t = \mathrm{A} + \mathrm{M} \left(\frac{1}{2}\bar{z} + \frac{3}{2}z_{t-1} - z_{t-2} \right) \ , \tag{F.11}$$

which can be rewritten, defining $w_t = z_{t-1}$, $\alpha = 1/2\,\bar{z}$, $\beta = 3/2$ and $\beta_2 = -1$, as a 4-dimensional system

$$\begin{aligned} \begin{pmatrix} z_t \\ w_t \end{pmatrix} &= \begin{pmatrix} \mathrm{A} + \mathrm{M}\,\alpha \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_1 \mathrm{M} & \beta_2 \mathrm{M} \\ I & 0 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ w_{t-1} \end{pmatrix} \\ s_t &= \mathrm{B} + \mathrm{N}\, s_{t-1} \ , \end{aligned} \tag{F.12}$$

whose stability depends on the eigenvalues of matrix N. Fig. F.21 depicts the eigenvalues of matrix N as function of the policy parameter ϕ_π. When $\phi_\pi = 1$, two complex eigenvalue

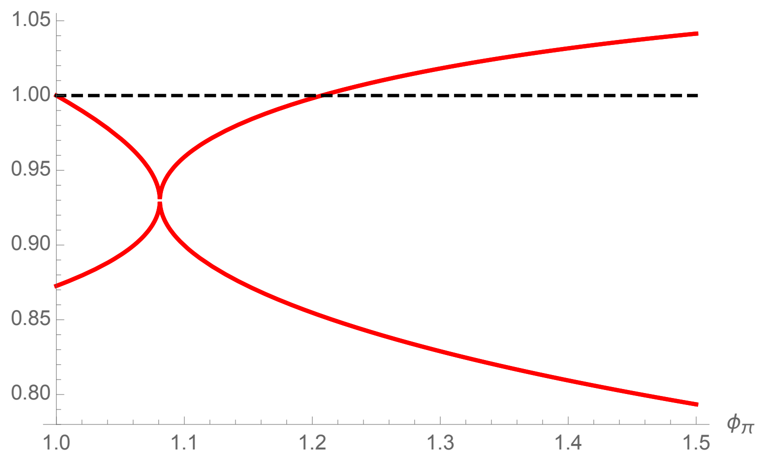


Figure F.21: Absolute value of eigenvalues of matrix N as function of ϕ_π.

are exactly on the unit circle, while the others are within the unit circle, meaning that there in $T1$ there is a continuum of stable equilibria under homogeneous LAA forecasts. As ϕ_π increases from 1 to 1.5, eigenvalues move from within to outside the unit circle. Therefore, under the policy regimes implemented in treatments $T2$ and $T3$, the system is stable with homogeneous anchoring and adjustment forecasting heuristics. On the contrary, under the policy regime of $T4$ the system exhibits explosive complex eigenvalues and it is therefore unstable. The intuition for this result is the following. Start from equilibrium and suppose there is a positive shock in inflation expectations. This will cause actual inflation to increase via the NKPC, but at the same time it will lower output via an higher interest rate. When the interest rate is aggressive enough, output fluctuations are large and they are further amplified by trend-extrapolating LAA rule. This has a negative impact on inflation, which can overshoot the target, leading the central bank to lower the interest rate reversing the trend in the output gap. The combination of strong interest rate reaction and trend extrapolation may lead small initial deviations from equilibrium to be amplified over time, causing oscillatory divergence. Simulations of system (F.10) with observable sample mean $\bar{z}_{t-1}$ confirm these results and are reported in Fig. F.22 for $\phi_\pi = 1.015$ and $\phi_\pi = 1.5$.[28] Notice that, in order to initialize system (F.10), we need to set the first two values z_1 and z_2. We fix the initial value at steady state, i.e. $(y_1, \pi_1)' = ((1-\rho)\bar{\pi}/\lambda, \bar{\pi})$, and we define $(y_2, \pi_2)'$ on a grid defined by points $y_2 = \{y_1 - 0.1, y_1, y_1 + 0.1\}$ and $\pi_2 = \{\pi_1 - 0.1, \pi_1, \pi_1 + 0.1\}$. Each line

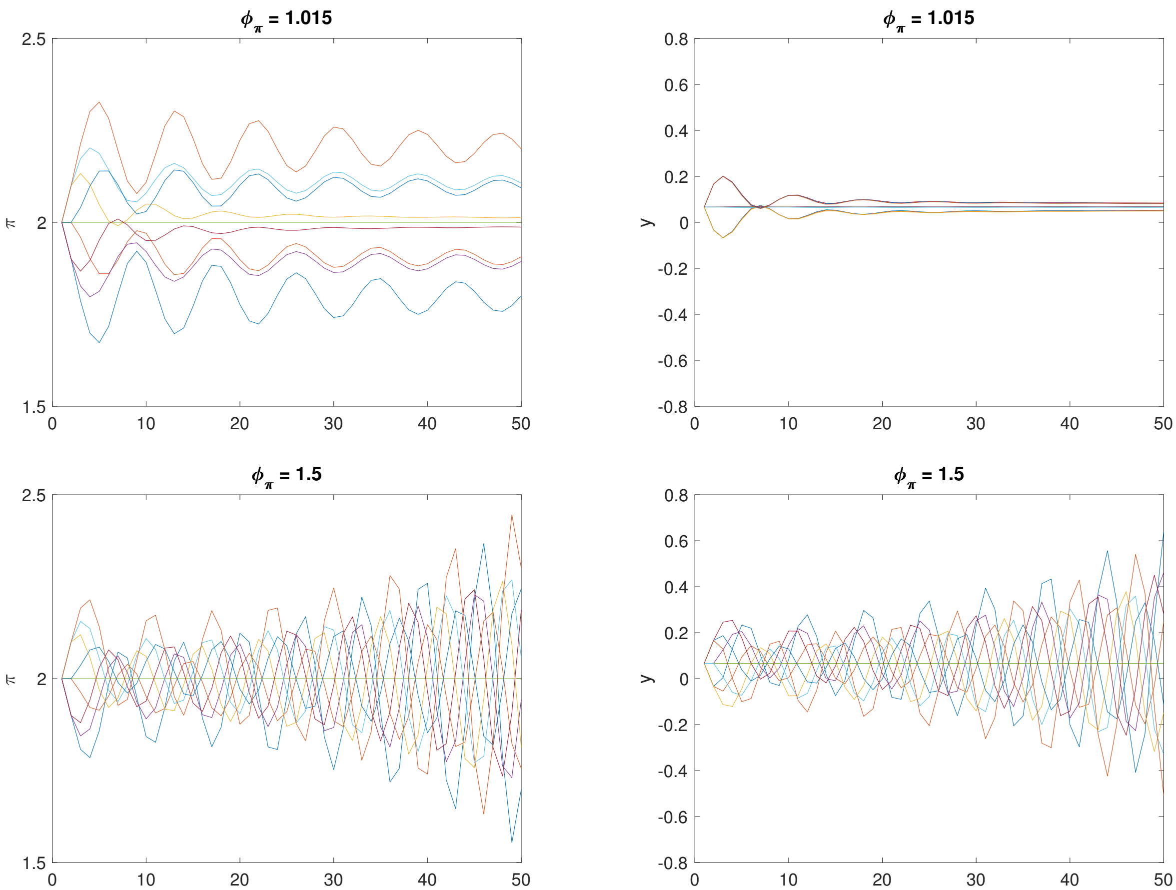


Figure F.22: Simulated dynamics of inflation (left panels) and output gap (right panels) under LAA heuristics for $\phi_\pi = 1.015$ (top panels) and $\phi_\pi = 1.5$ (bottom panels).

corresponds to simulated dynamics for different initial conditions. When monetary policy is not too aggressive the system is stable, although convergence can be very slow due to an eigenvalue almost on the unit circle. On the other hand, when the policy reaction is strong,

[28] Dynamics for $\phi_\pi = 1.005$ are similar to those obtained for $\phi_\pi = 1.015$

the system is unstable displaying oscillatory divergence.

Appendix G. One-step-ahead simulations for all groups

In this section we report the results of one-step ahead predictions for all experimental economies. Left panels in Figs. G.23–G.34 display experimental data together with the one-step-ahead predictions under the HSM, while right panels depict the evolution over time of the weights of the four considered heuristics.

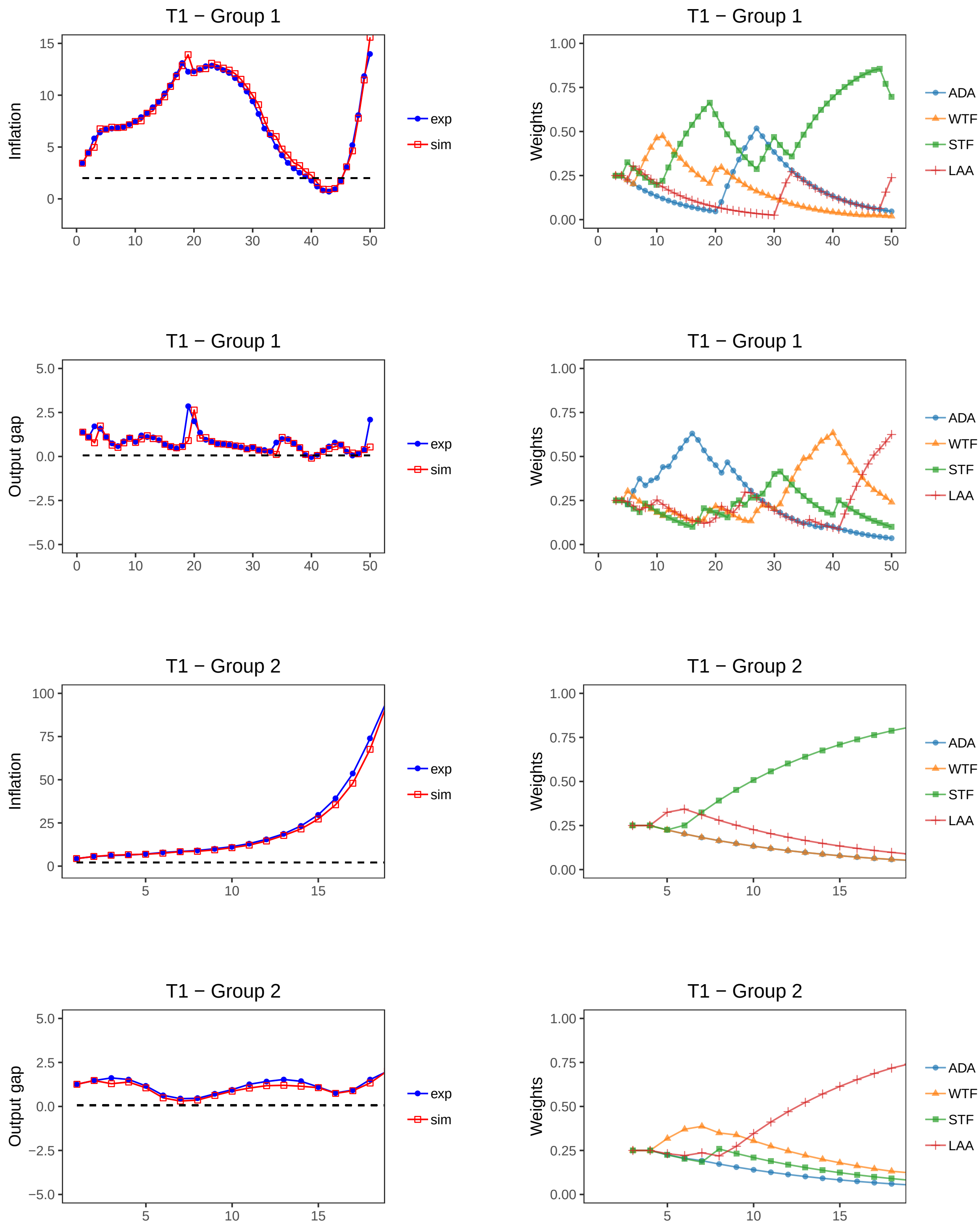


Figure G.23: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T1$ (groups 1–2).

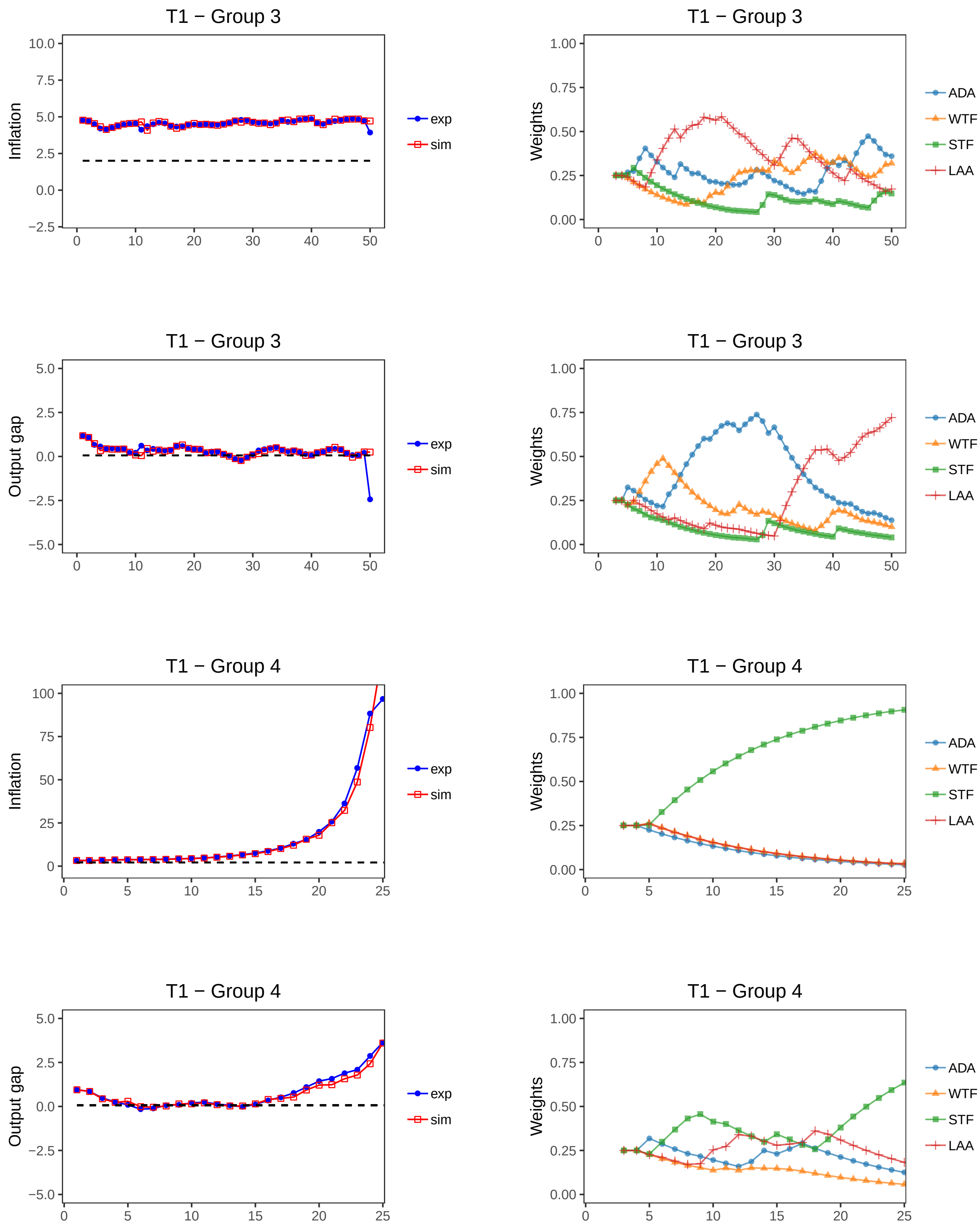


Figure G.24: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T1$ (groups 3–4).

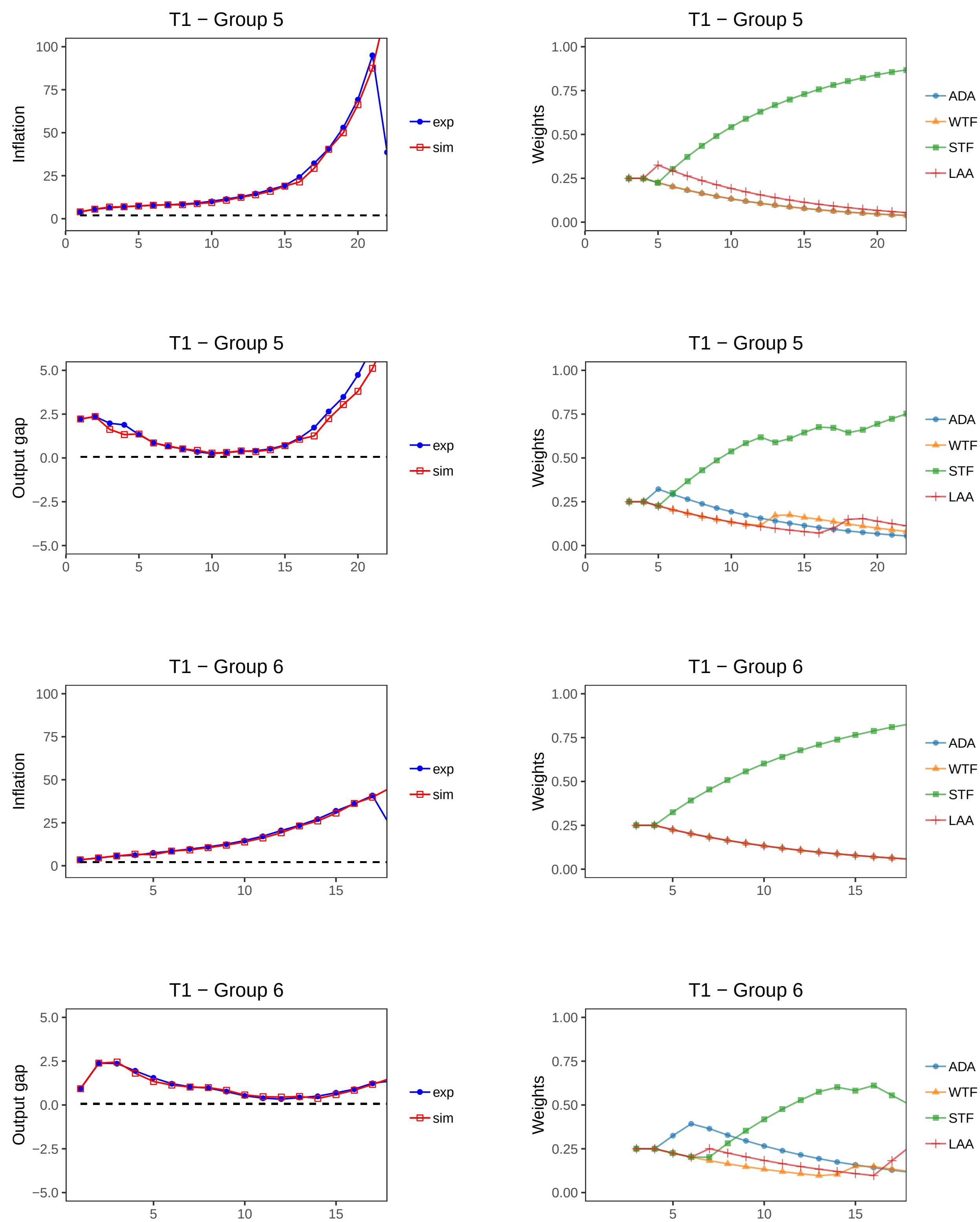


Figure G.25: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T1$ (groups 5–6).

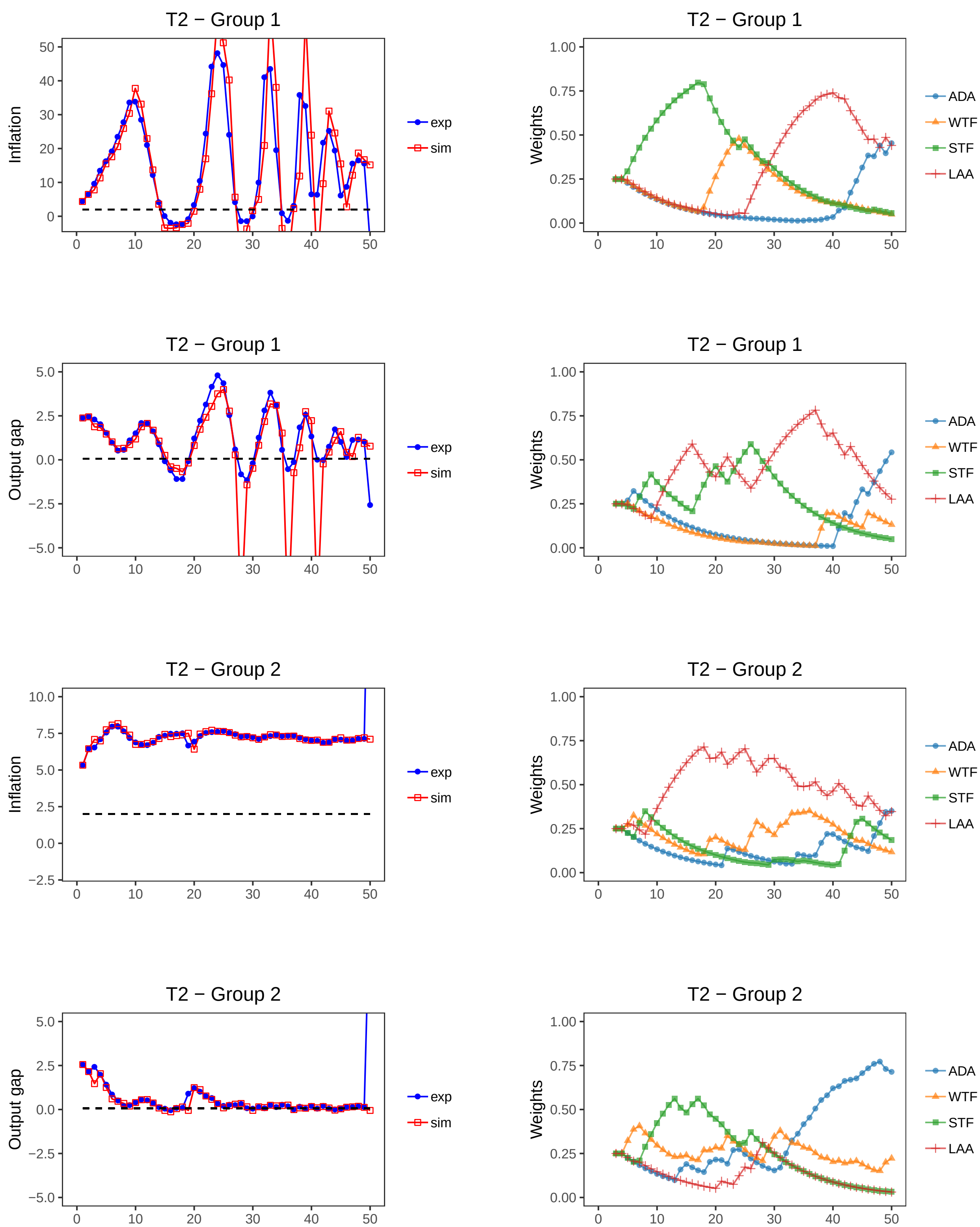


Figure G.26: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T2$ (groups 1–2).

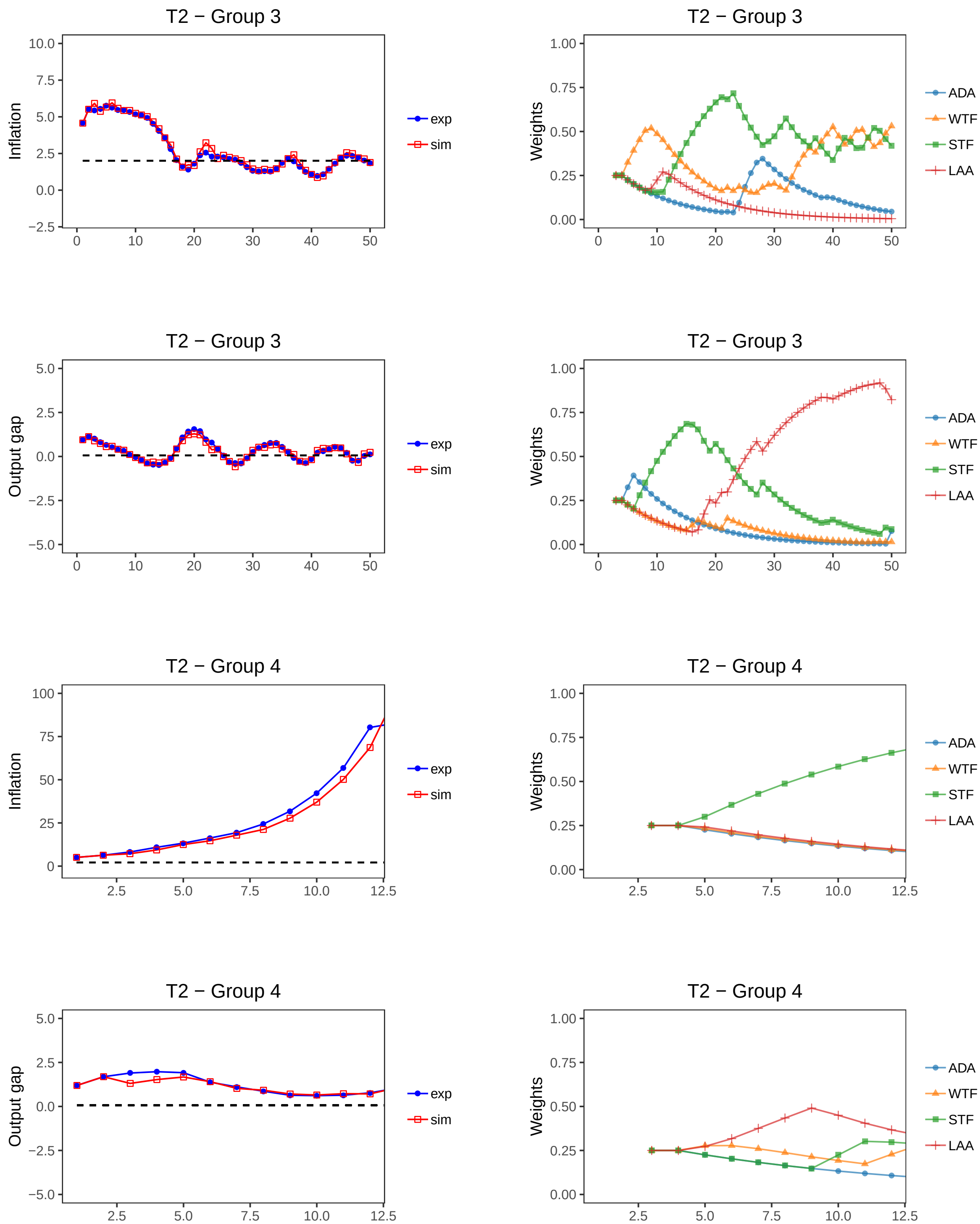


Figure G.27: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T2$ (groups 3–4).

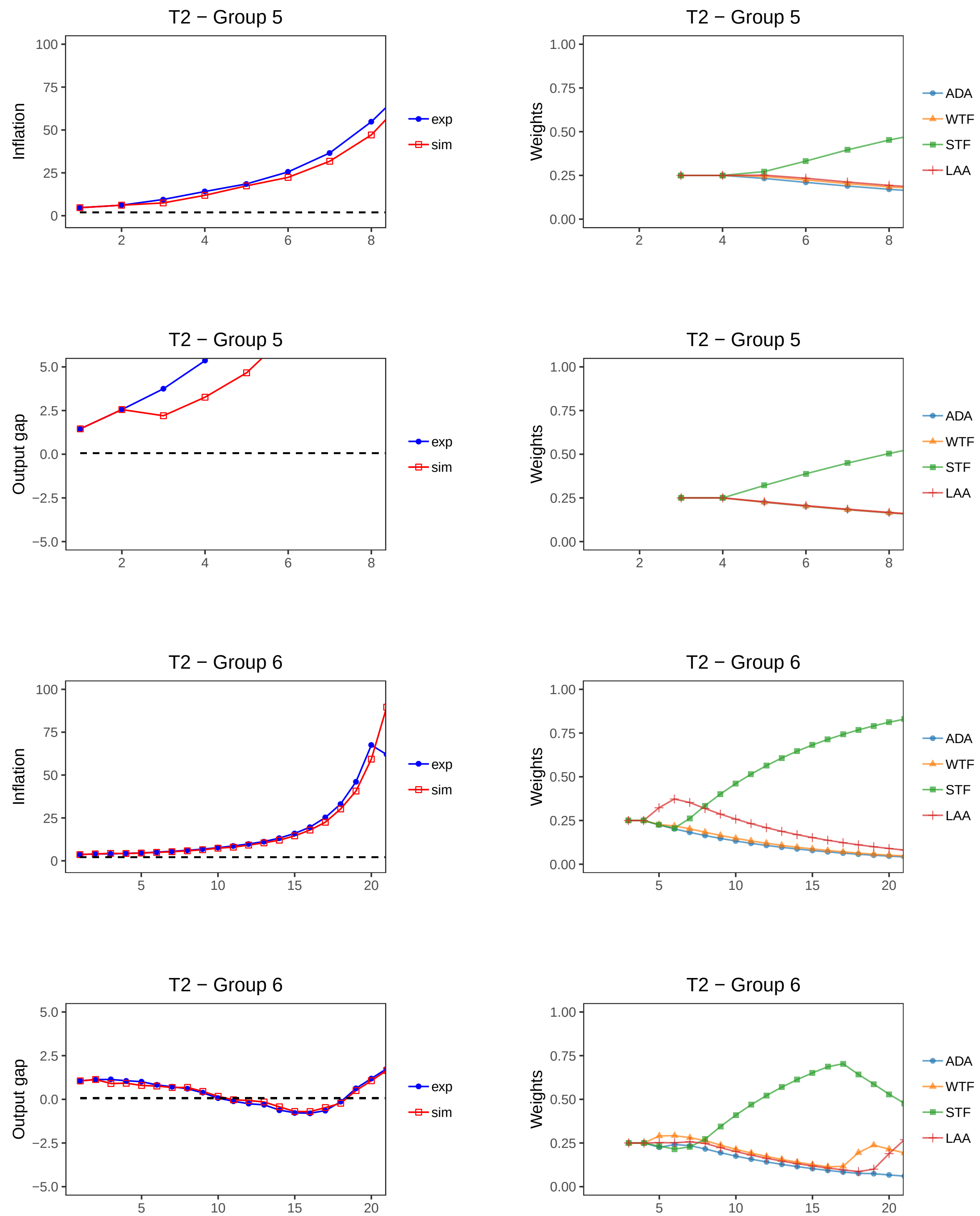


Figure G.28: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T2$ (groups 5–6).

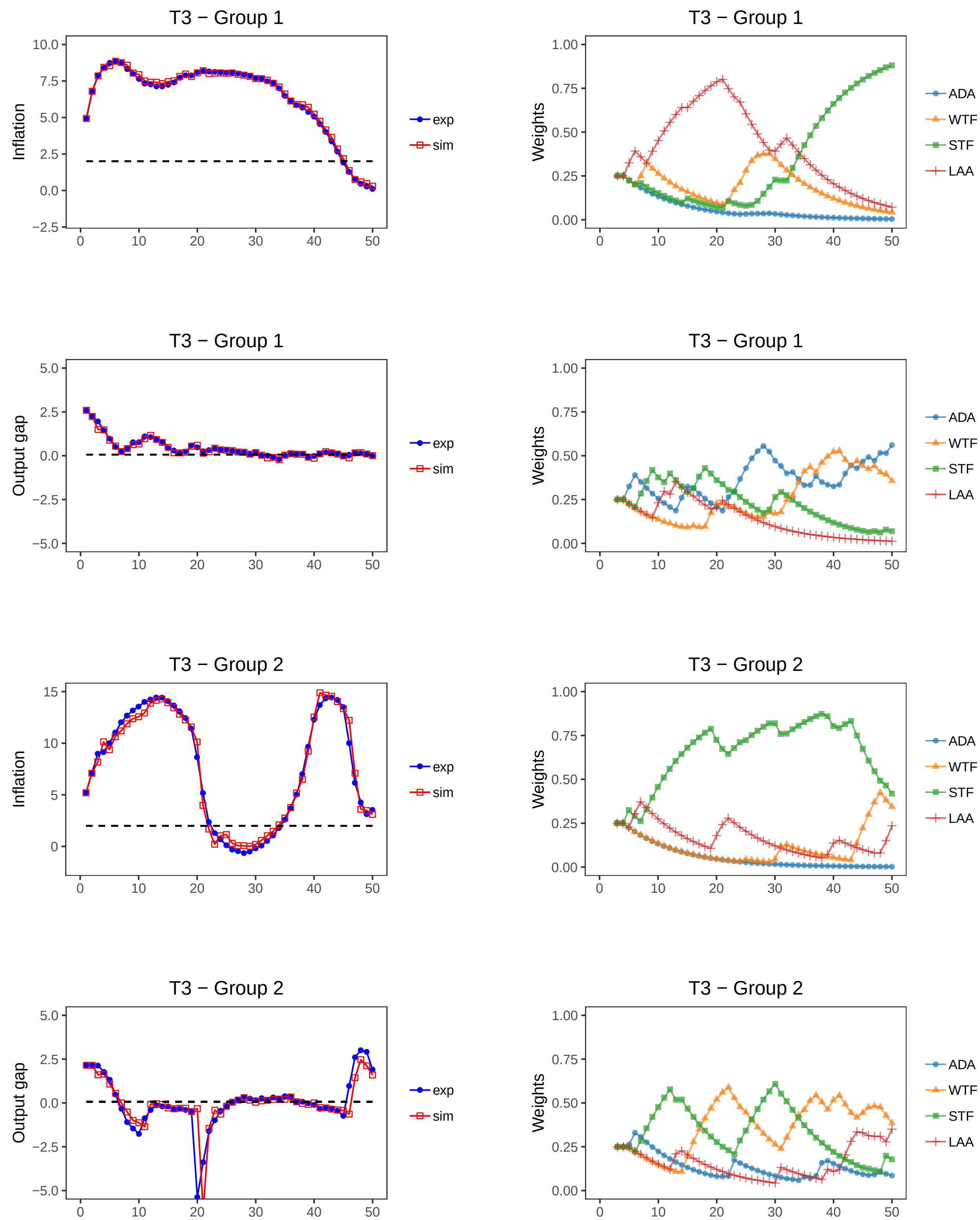


Figure G.29: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T3$ (groups 1–2).

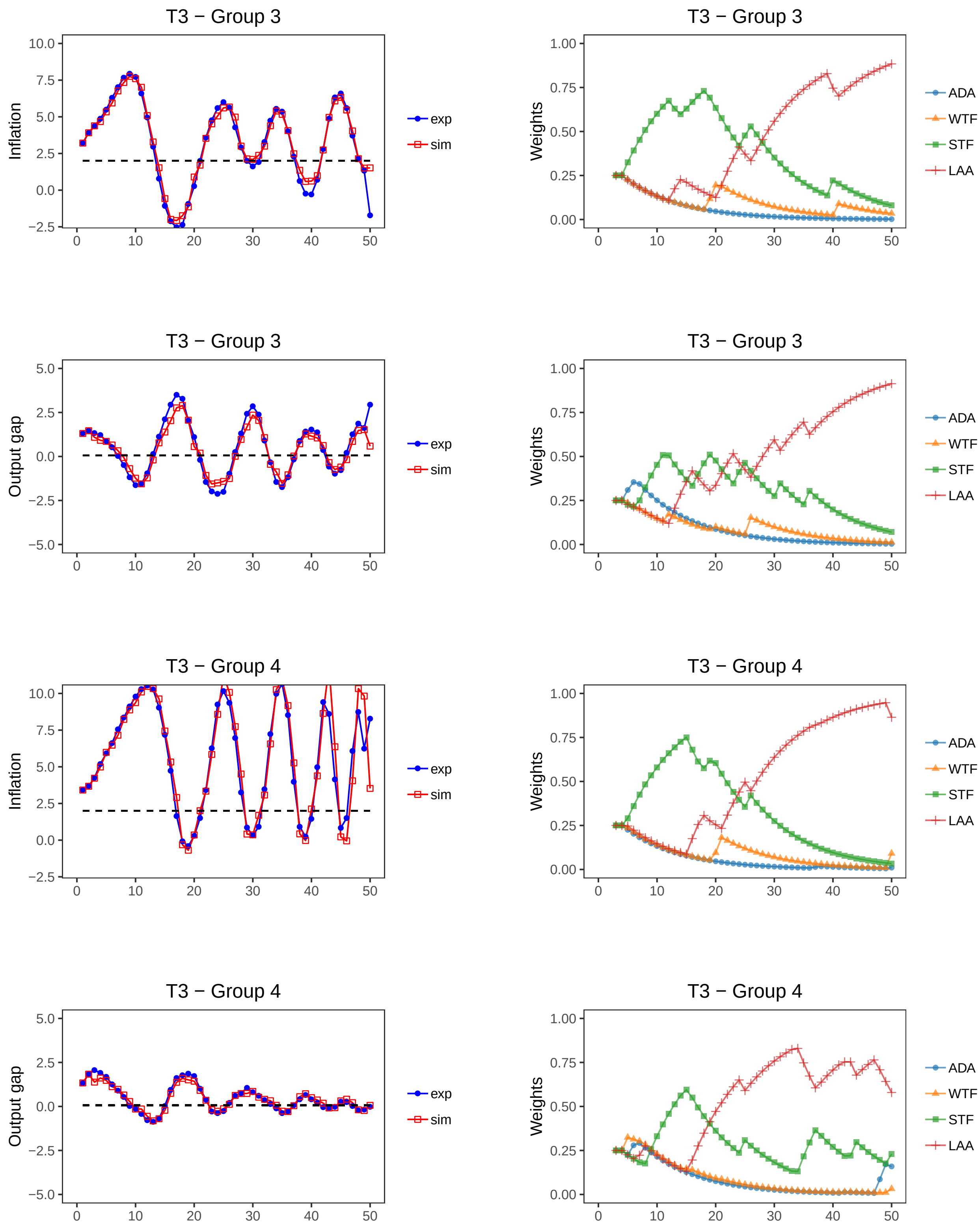


Figure G.30: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T3$ (groups 3–4).

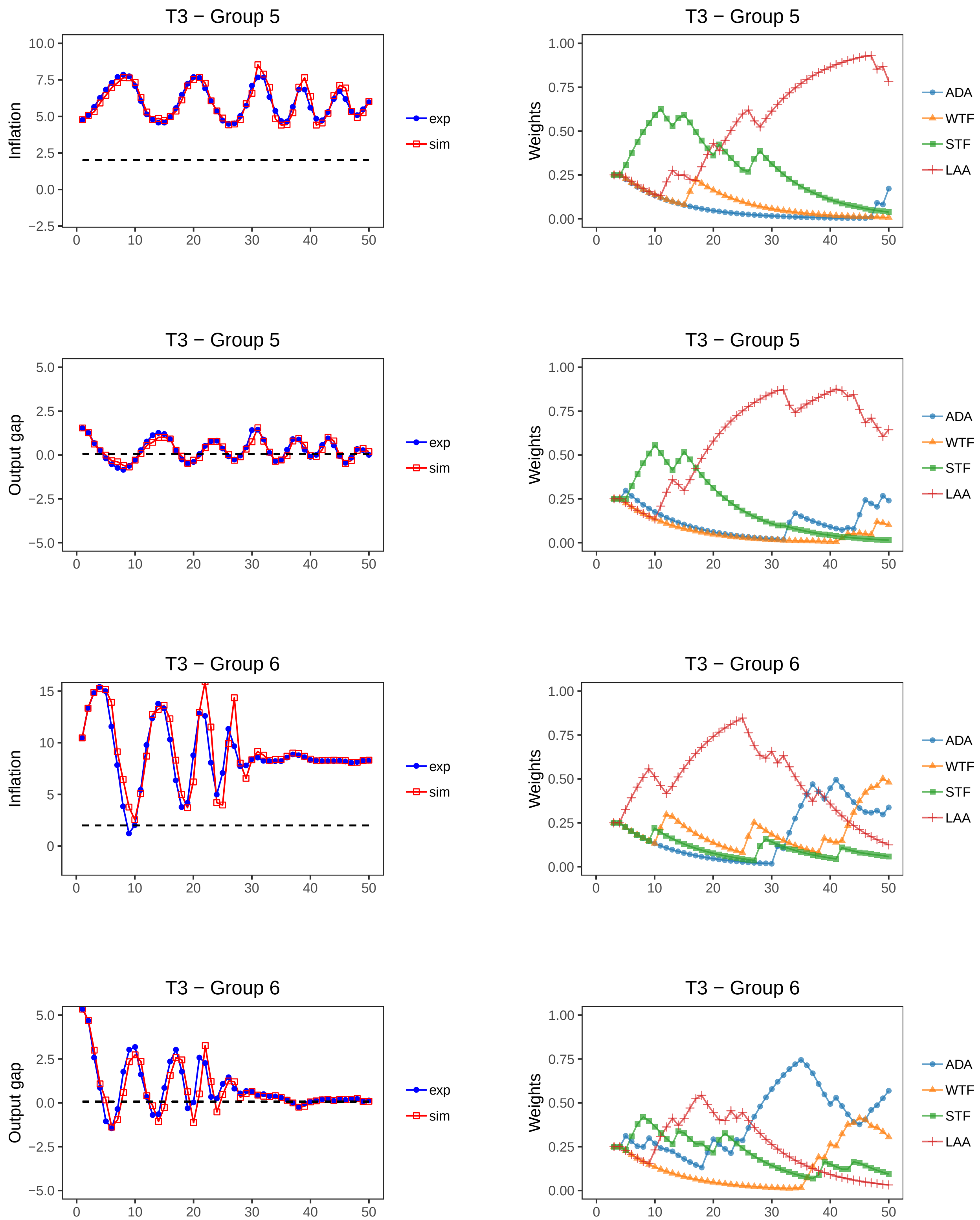


Figure G.31: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T3$ (groups 5–6).

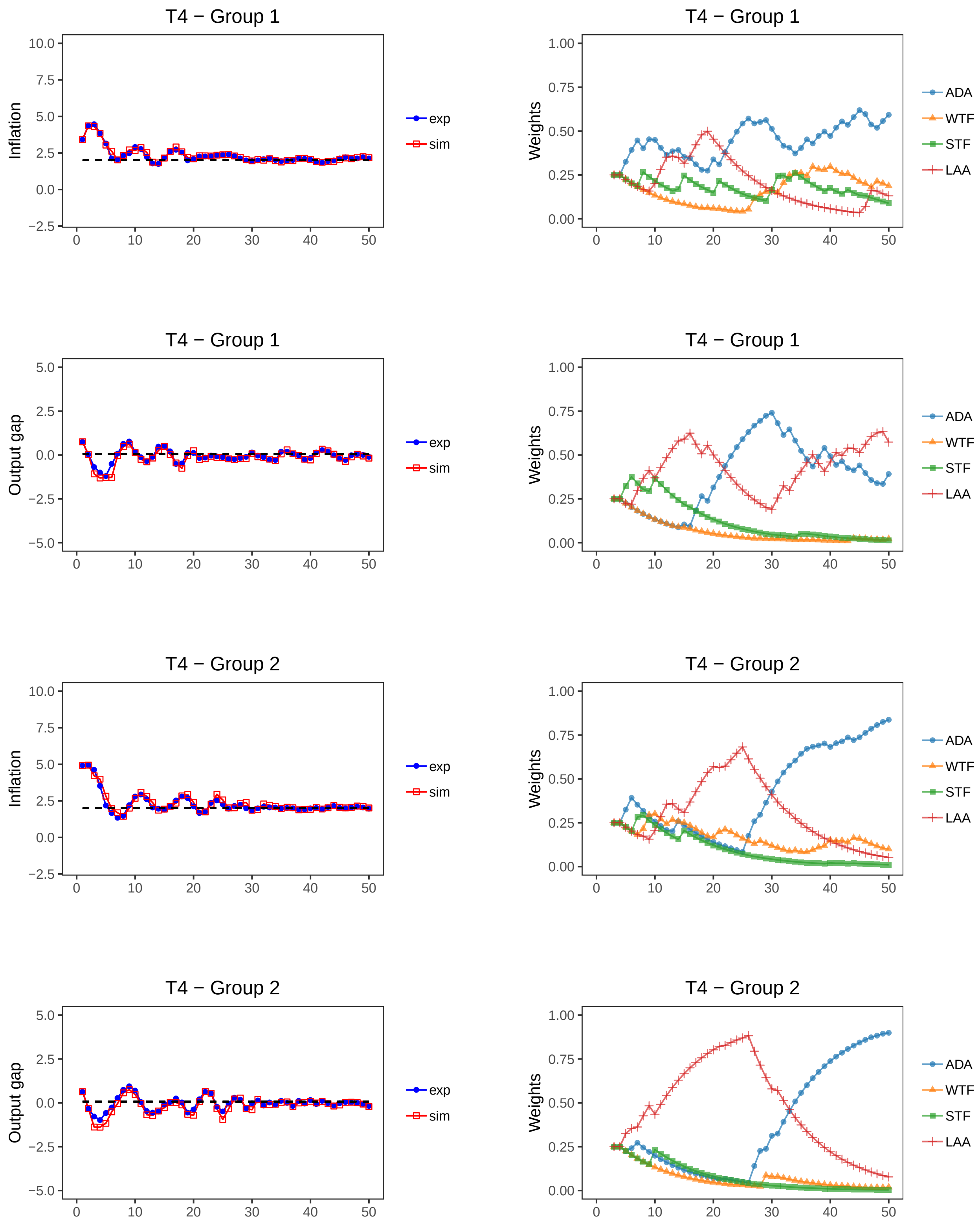


Figure G.32: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T4$ (groups 1–2).

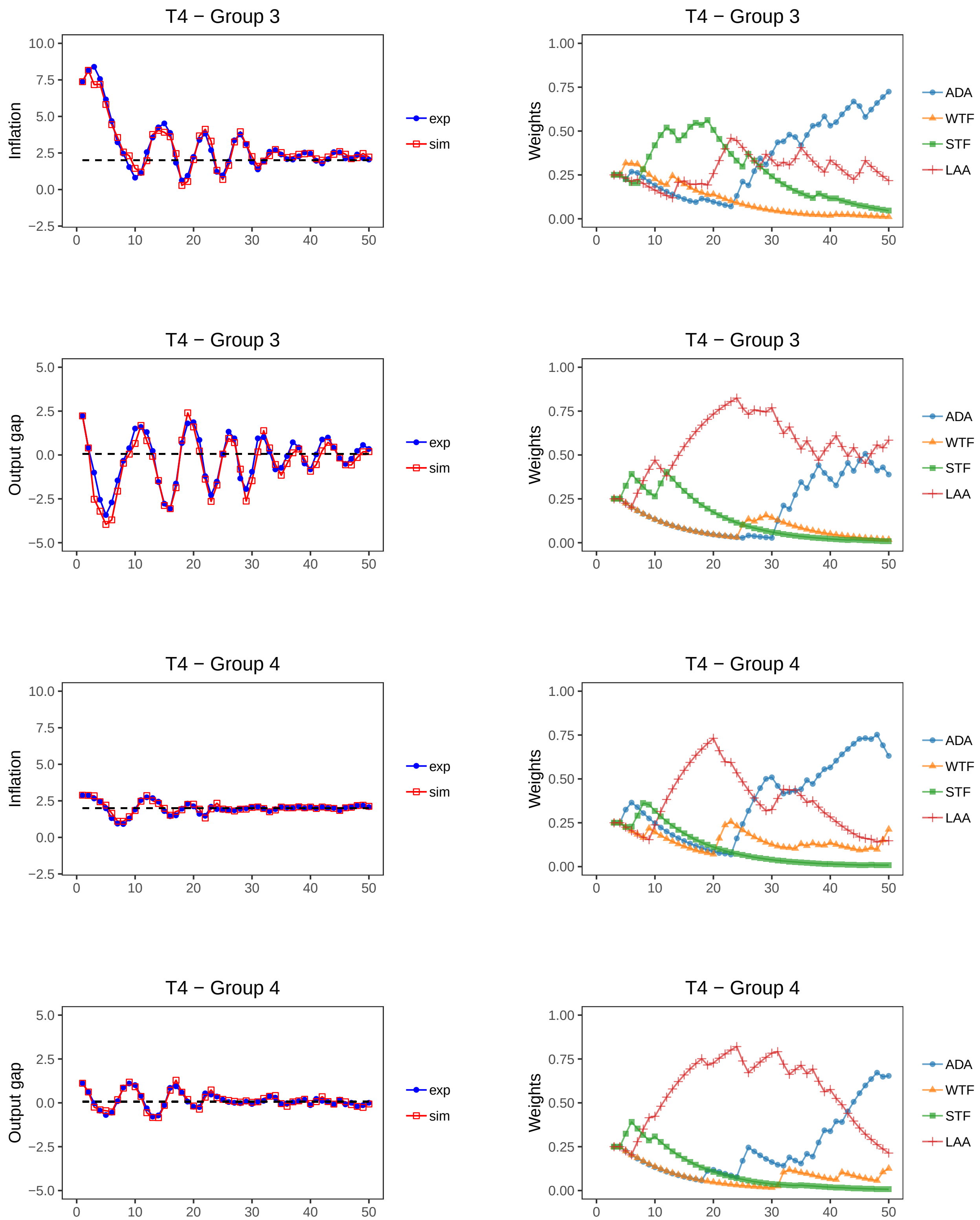


Figure G.33: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T4$ (groups 3–4).

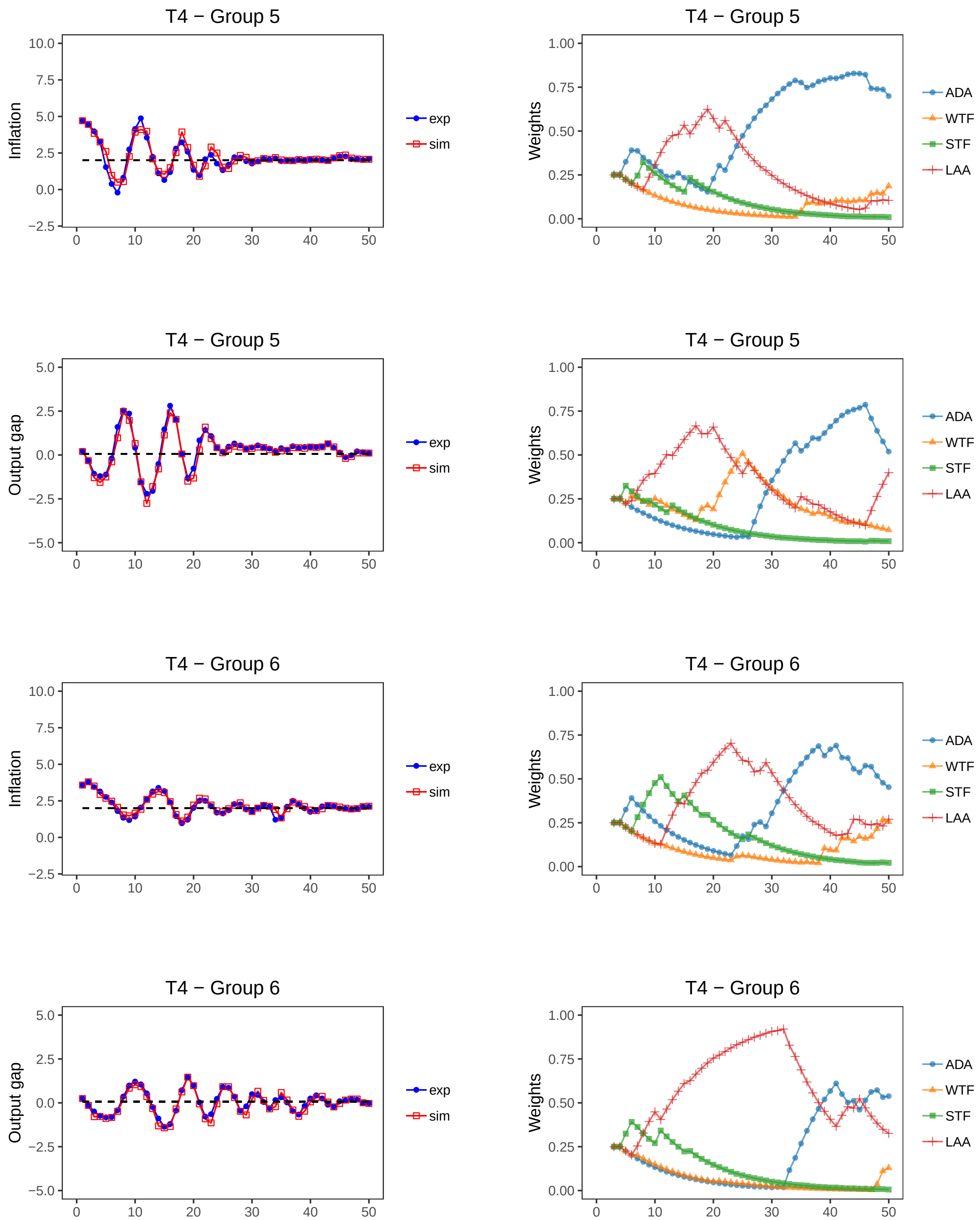


Figure G.34: Realized and simulated inflation and output gap (left panels) with corresponding simulated weights of 4 heuristics for $T4$ (groups 5–6).