BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2022-33

Last updated: July 15, 2022

A Horse Race of Monetary Policy Regimes: An Experimental Investigation

by Olena Kostyshyna,[1] Luba Petersen[2] and Jing Yang[1]

[1]Canadian Economic Analysis Department
Bank of Canada
kost@bankofcanada.ca, jyang@bankofcanada.ca

[2]Department of Economics
Simon Fraser University
lubap@sfu.ca

DOI: https://doi.org/10.34989/swp-2022-33 | **ISSN** 1701-9397

Acknowledgements

We have benefited from useful discussions with Cars Hommes, Rhys Mendes, Yang Zhang, Oleksiy Kryvtsov, Athanasios Orphanides, Chen Lian, Benjamin Schoefer, Jón Steinsson, Emi Nakamura, Ryan Chahour, Alexander Ueberfeldt, Isabelle Salle, Joel Wagner, Herbert Dawid, and Tomasz Makarewicz as well as seminar and conference participants at the Bank of Canada, Bielefeld University, C.D. Howe Institute, Florida State University, University College Dublin, University of California (UC) Berkeley, the Max Bell School of Public Policy's workshop on the Bank of Canada's inflation target renewal, the 27th International Conference of the Society for Computational Economics, the 11th Workshop on Theoretical and Experimental Macroeconomics, the 3rd Behavioral Macroeconomics Workshop at the University of Bamberg, and the Atlantic Canada Economics Association Annual Meeting. Finally, we thank Ekeme Afia, Michael Mirdamadi, Fatemeh Mokhtarzadeh, and Ryan Rholes for outstanding research assistance and Justin Baize for excellent programming support.

Abstract

We provide a comprehensive assessment of five monetary policy regimes—inflation targeting (IT), dual mandate (DM), average inflation targeting under 4-period (AIT-4) and 10-period (AIT-10) horizons, price level targeting (PLT), and nominal GDP level targeting (NGDP)—in a unified experimental framework. We study how participants can understand different regimes and form expectations during periods of economic stability and during a demand-driven recession that temporarily brings the economy to the ELB. Our results suggest a distinct ranking of policy regimes in terms of their ability to achieve macroeconomic stability. DM and IT are the most stabilizing regimes, followed by AIT and then level-targeting regimes PLT and NGDP. AIT with a shorter horizon performs better than AIT with a longer horizon. Monetary policy regimes that are framed around the inflation rate (e.g., AIT-10) are found to deliver significantly more stable economic outcomes than those that target price levels (PLT). Participants have greater difficulty understanding regimes that are more history-dependent and forecasting in the rationally expected direction. They instead rely on trend-chasing heuristics to form their expectations. Trend-chasing forecasting is more destabilizing for regimes with more history dependence. Participants also "need to see it to believe it." PLT and NGDP initially have mixed success at achieving their targets, and these regimes do not gain credibility before the economy enters into the ELB, where credibility is needed more than ever.

Topics: Inflation targets; Monetary policy; Monetary policy framework; Monetary policy communications
JEL codes: C9, D84, E52, E58

1 Introduction

Monetary policy has evolved significantly in the aftermath of the Global Financial Crisis and the pandemic. Leading central banks, explicitly following an inflation-targeting framework for decades, have driven their policy rates to their effective lower bounds in an effort to provide liquidity to markets and stimulate sluggishly low inflation expectations and economic growth. With limited traction through their actual policy rates, many central banks have begun to entertain monetary frameworks that promise to keep interest rates "lower for longer" in an effort to bolster inflation expectations in the short run. For example, an average inflation targeting (AIT) framework would involve a central bank making up for recent past shortfalls in inflation by temporarily accepting above-target inflation as the economy recovers.

The advantages of level-targeting mandates have recently gained attention with suggestions of price-level targeting (PLT) (Evans [2012], Williams [2017], Bernanke [2017], Bullard [2018]) and nominal GDP level targeting (NGDP) (Carney [2012]) as alternative policy mandates. These history-dependent regimes can be more powerful in stabilizing the economy as past misses must be made up (ITR [2011], Carney [2012]). While PLT has gained some popularity in recent monetary policy discussions, there is very limited experience with it: PLT was briefly implemented only in Sweden in the 1930s (Jonung [1979], Berg and Jonung [1999]). The United States defined its monetary policy framework as flexible average inflation targeting only recently in August 2020 (Powell), and nominal GDP level targeting has not been implemented anywhere yet.

The assessment of PLT in the 2011 Inflation Control-Target Renewal in Canada (ITR [2011]) suggests modest but economically significant benefits from PLT and even higher potential gains associated with this regime at the zero lower bound. However, the superior performance of PLT depends on agents' forward-looking expectations and credibility of the regime (ITR [2011]). Any advantages in the performance of such a history-dependent regime could diminish or even reverse if the expectations channel does not work well, i.e. people do not understand how policy regime works, or if policy regime is not credible (ITR [2011], Carney [2012]). "People must generally understand what the central bank is doing - an admittedly high bar" (Carney [2012]).

Evidence about people's understanding of these alternative monetary policy regimes is very limited. Coibion et al. [2020a] show that surveyed households in the United States have difficulty understanding the newly introduced AIT framework and distinguishing it from IT regime. Hoffmann et al. [2021] provide results that are more encouraging for AIT where German households revise their expectations consistently with AIT policy. A key difference between these two papers is that Coibion et al. [2020a] study actual shift to AIT in the U.S., while Hoffmann et al. [2021] explore *hypothetical* policy change by ECB.

In this paper we provide a comprehensive assessment of five monetary policy regimes – IT, dual mandate, AIT, PLT, and NGDP – in a unified experimental framework. This approach allows us to study how effectively people can understand and form macroeconomic expectations under the competing policy frameworks both during periods of economic stability away from the ELB and during demand-driven recessions at the ELB. Laboratory experiments have been used extensively in the literature to compare inflation targeting with price-level targeting (Hommes and Makarewicz [2021a], Amano et al. [2011], Arifovic and Petersen [2017], Salle [2021]), a dual mandate (Cornand and M'baye [2018], Hommes et al. [2019b]), and average inflation targeting (Salle [2021]).

To this end, we take the simple New Keynesian model used in the Bank of Canada's own theoretical analysis to the laboratory and run an experimental horse race of the various mandates under consideration. Using a standard learning-to-forecast experimental setup, we incentivize groups of participants to form expectations about future inflation and output gap. The aggregated expectations, combined with shocks to the natural rate of interest, are fed into the economies' data-generating process and drive macroeconomic dynamics. In a between-subject design, we systematically compare aggregate dynamics and expectation formation under the different monetary policy regimes. We further explore the robustness of the competing regimes in periods of stability away from the effective lower bound (ELB) and in response to significant negative demand shocks that temporarily drive economies to the ELB. The economies are parameterized such that, under rational expectations, NGDP and PLT should have a small advantage over AIT-10, AIT-4, DM, and IT.

Our experiments suggest a distinct ranking of the monetary policy regimes in terms of their ability to achieve macroeconomic stability. Rate-targeting regimes such as IT, DM, and AIT significantly outperform level-targeting regimes such as PLT and NGDP in terms of their ability to minimize deviations of inflation, output gap, and nominal interest rates from target. IT, DM, and AIT with a short horizon of four quarters deliver very similar performances. The IT regime is among the frontrunners in this monetary policy horse race despite being the predicted loser under rational expectations.

One of our key conclusions is that the length of horizon in the monetary policy rule matters. Greater history dependence in a mandate results in less-stable economies. We find that AIT with a shorter horizon (4-quarter) performs better than AIT with a longer horizon (10-quarter). This observation is similar to Amano et al. [2020], who find that a short horizon is optimal in the presence of backward-looking expectations. Indeed, the backward-looking expectations observed in all our treatments generate considerable inflation volatility. Mandates such as AIT-10, PLT, and NGDP require the central bank to react to past economic deviations that may no longer be relevant for the current economic situation. This can further unanchor participants' expectations.

Framing also matters. We find that highly history-dependent inflation targeting mandates such as AIT-10 deliver significantly more stable economic outcomes than PLT. The framing of AIT in terms of the inflation rate presents less cognitive burden for inflation forecasters than the framing of PLT in terms of deviations of the price level from target. It is more challenging for participants to interpret what deviations of price level from target mean for future inflation.

Our individual-level analysis further suggests that participants have difficulty understanding the basic 'directionality' associated with the various monetary policy regimes, i.e. forecasting in the rationally expected direction. This is especially evident in the level-targeting regimes. But even participants who do forecast in the correct direction do not fully internalize the stabilizing effects of monetary policy, i.e. they react too little.

Finally, credibility is more challenging to maintain in level-targeting mandates. Broadly speaking, participants 'need to see it to believe it.' While inflation regularly returns to target in the rate-targeting treatments, it can take quite a long time for price and nominal GDP levels to return to target in the level-targeting treatments. Participants in the PLT and NGDP treatments eventually grow skeptical about the central bank's ability to achieve their targets, especially as their economies enter into the ELB and the central bank alone can no longer stimulate the economy back to its intended targets.

Our paper is organized as follows. Section 2 presents related literature. Section 3 lays out our experimental design and hypotheses. Section 4 presents our experimental findings, and Section 5 concludes.

2 Literature

Our paper contributes to an extensive discussion of the design of monetary policy to manage expectations. We explore five different monetary policy mandates, some of which have already been compared in the literature (Arifovic and Petersen [2017], Ahrens et al. [2017], Cornand and M'baye [2018], Assenza et al. [2019], Hommes et al. [2019a], Kryvtsov and Petersen [2021], Salle [2021], and others). Our experiment is the first to pursue a dual mandate (DM) of equal weight on both inflation and the output gap, and furthermore, it's one of the strongest studied in the literature. Cornand and M'baye [2018] and Hommes et al. [2019b] show that a flexible inflation targeting (IT) mandate, whereby the central bank reacts to both inflation and the output gap, outperforms a strict IT mandate in achieving greater inflation stability. However, in their experiments, the reaction coefficient on the output gap is relatively small compared to inflation ($\phi_\pi = 0.5$). We study two largely unexplored mandates in our monetary policy horse race – average inflation targeting (AIT) and nominal GDP targeting (NGDP). Like PLT, both AIT and NGDP are history-dependent mandates that require the central bank to make up for past deviations from target.

Other studies have investigated price-level targeting (PLT) relative to inflation targeting. In one of the earliest experiments, Amano et al. [2011] study expectation formation in an exogenously evolving environment, unaffected by participants' expectations. They observe that IT outperforms PLT. While subjects' forecasts adjust in the intended direction under PLT, expectations are not sufficiently rational, showing lack of understanding of this monetary policy regime. Arifovic and Petersen [2017] compare IT and PLT with qualitative and quantitative communication of the central bank's target price level. In their experiment, the authors reframe the PLT as an IT with evolving inflation target to simplify the forecasting tasks for participants and announce it either quantitatively or qualitatively ("the central bank is aiming for higher inflation"). Similar to us, Arifovic and Petersen [2017] study the mandates both away from the ELB and at the ELB after the economy experiences a significant negative and persistent demand shock. They find that neither form of communication of the price level target works effectively to stabilize economy and expectations at the ELB, and there is little to be gained by using PLT relative to IT.

More recently, Hommes and Makarewicz [2021b] show that an explicit PLT can be stabilizing only if it is sufficiently responsive to both the output gap and prices. In a concurrently run experiment, Salle [2021] finds that PLT outperforms IT. Salle's framework is a simple neo-Fisherian model with flexible prices in which subjects forecast only one variable: inflation. The task of our subjects of forecasting both inflation and output can be more challenging, and it is further complicated by encountering a very large ELB shock. Hommes and Makarewicz [2021a] use a non-linear New-Keynesian model where subjects forecast both inflation and output, and they find that only very strong responses in PLT can deliver stable results. We find that while PLT can be relatively stable during the pre-shock period, it results in explosive dynamics with unraveling deflation following ELB shock. In Hommes and Makarewicz [2021a], economies experience a brief

ELB episode with deflation because of non-persistent iid shocks. The more severe and longer episode at the ELB likely contributed to the instability observed in our PLT experiment.

Our AIT results complement those of Salle [2021], who investigates expectation formation under three different types of AIT policy rules: average inflation calculated over two or four periods and a two-period lagged average inflation with a stronger central bank reaction coefficient. The design of our AIT policy treatments differ in that we implement policy rules with four- and 10-period lags of inflation. Similarly to our results, Salle [2021] finds that participants have greater difficulty coordinating their expectations on the central bank's inflation target when the central bank reacts to more lags of inflation in their average inflation calculation. She also finds that having a more aggressive policy reaction coefficient works better to manage expectations.

Evidence about the public's understanding of AIT is very limited. Coibion et al. [2020a] show that U.S. households have difficulty understanding the newly introduced AIT framework of the Federal Reserve. Comprehension of AIT and elicited inflation expectations differed little between those who were aware of the Federal Reserve's announcement about AIT and those who were unaware. Furthermore, in a randomized control trial (RCT), people did not respond differently to information about the Federal Reserve's average inflation target than those who received information about its existing inflation target. This suggests that the expectations channel on which AIT relies on for its superior performance relative to IT may not work as intended. Hoffmann et al. [2021] provide results that are more encouraging for AIT, where German households revise their expectations consistently with AIT policy. In their RCT, respondents were provided with information about a hypothetical alternative ECB monetary policy regime akin to AIT used by the Federal Reserve. They find that inflation expectations increase for those in the treatment group. With additional information about current inflation, respondents update their inflation expectations path consistently with monetary policy intentions under AIT. The results in these two papers could be different because Coibion et al. [2020a] study actual shift to AIT in the U.S., while Hoffmann et al. [2021] explore *hypothetical* policy change by ECB. Hoffmann et al. [2021] also note that the distribution of inflation expectations in their survey is close to actual inflation outcomes, suggesting that respondents in Germany follow inflation quite carefully.

Our work also contributes to the literature studying monetary policy at the ELB. Arifovic and Petersen [2017], Ahrens et al. [2017], Hommes et al. [2019a], Assenza et al. [2019], and Kryvtsov and Petersen [2021] show consistent evidence that economies can become highly unstable at the ELB and deflationary episodes can occur. Deflationary spirals can develop under different monetary policy regimes: IT and state-dependent IT in Arifovic and Petersen [2017], Ahrens et al. [2017], Assenza et al. [2019]; IT and PLT in Hommes et al. [2019a], Kryvtsov and Petersen [2021]). In our work, we observe deflationary spirals in history-dependent monetary policy regimes. Some of the studies show that deflationary episodes can sometimes, but not always, be tempered and economic stability be restored through fiscal stimulus and credible central bank projections.

We also contribute to the literature on understanding of monetary policy and on expectation formation in macroeconomic laboratory experiments. Learning-to-forecast experiments (LtFEs) have been employed to understand how individuals coordinate their forecasts in macroeconomic economies largely driven by aggregate expectations (Adam [2007]). LtFEs have demonstrated how the nature of inflation-targeting monetary policies can influence the nature of expectation formation and macroeconomic stability (Pfajfar and Žakelj [2014], Kryvtsov and Petersen [2013]). Assenza et al. [2019] show that heterogeneous expectations tend to self-organize on different forecasting rules depending on monetary policy. Subjects are generally found to form non-rational expectations in that they use historical information, rather than relevant current shocks and the data-generating process including monetary policy rule, to formulate their forecasts. More aggressive reaction coefficients on inflation and output gap in the central bank's Taylor rule encourage more stable forecasting behaviour and aggregate stability. Macroeconomic experiments have been used to identify the heuristics individuals and groups use to forecast. Anufriev et al. [2013] show that the stability of a system depends on the composition of forecasting rules. Our experiment highlights the relative importance of trend-chasing and constant gain heuristics in driving participants' forecasting behavior. Our analysis highlights how the degree of trend-chasing increases as the policy mandate becomes more history-dependant.

Our paper is also related to the literature that studies the dynamics of macroeconomic models and stabilization properties of monetary policy regimes when agents form non-rational expectations (Mitra [2003], Honkapohja and Mitra [2014], Honkapohja and Mitra [2020]). Mitra [2003] showed that NGDP level targeting can be stable under learning by non-rational agents for all model parameterizations when nominal interest is set based on agents' expectations. Honkapohja and Mitra [2014] and Honkapohja and Mitra [2020] show that properties of NGDP-level targeting and PLT do not hold when agents form expectations adaptively. If agents learn about forecasting inflation and output from historical data (as we find experimental participants do), then these history-dependent regimes do not perform better than IT. Both of these regimes require additional guidance about the path of targeted variables to improve their performance. Honkapohja and Mitra [2020] show that performance of PLT depends on the communication about the path of price target and credibility of the PLT regime. However, as discussed above, Arifovic and Petersen [2017] showed that communicating an evolving inflation target to achieve price level target proves confusing to the subjects, and as a result, this regime is less stabilizing than IT.

Our finding that AIT with shorter horizon performs better than AIT with longer horizon is similar to that in Amano et al. [2020]. They find that AIT with relatively short horizon is optimal in a two-agent New Keynesian model with a fraction of firms forming backward-looking expectations, and that in this case the properties of the economy are quantitatively similar to those under a price-level target. The latter conclusion is, however, in contrast with our experimental results, where PLT is much more unstable than AIT of both short and long horizon. Thus, experimental evidence suggests that AIT of shorter horizon is more likely to stabilize the economy than PLT.

3 Experimental Design

How do different monetary policy regimes perform in the experimental setting with human participants? Are participants able to understand these regimes? How do people form forecasts about the economy under different monetary policy mandates? Do these beliefs internalize different policy rules? To answer these questions, we design a laboratory economy where participants are given the opportunity to experience economies with different monetary policy regimes and to formulate beliefs.

Laboratory experiments are a useful tool in the policymaker's toolbox. Its advantages have been discussed in the literature (Duffy [2016]) and are as follows. First, we can explore alternative frameworks in the laboratory setting, which is not possible to do in practice. Second, in experiments, we do not impose assumption about rational expectations or any specific mechanism of expectations formation. In the world of wilderness of bounded rationality (Sargent [1999]), different choices of expectations formation can lead to different conclusions about the stabilization properties of monetary policy regimes (Evans et al. [2001], Woodford [2010]). In the laboratory, while we do not make such assumptions about expectation formation and leave it to the subjects to figure out the environments, we can study ex post how they form their expectations (similarly to Assenza et al. [2019] and Anufriev et al. [2013]) and whether different monetary policy regimes induce different types of expectations. We can learn about how much people understand each monetary policy in a very simple experimental environment and whether they form their expectations consistently with each policy regime. Experimental method is a very valuable tool to assess whether an expectations channel works in history-dependent regimes. Third, the evidence about learnability of outcomes in the experimental setting can serve as a selection criterion among alternative monetary policy frameworks similarly to how learnability of equilibrium can serve as a selection criterion in case of multiple equilibria (Sargent [1993], Evans and Honkapohja [2001]). If a certain policy regime is unable to stabilize a simple economy in the laboratory, it is not likely to perform well in a real economy, and/or it may present substantial communication and implementation challenges. Various challenges associated with price-level targeting were discussed in ITR [2011].

3.1 Data-generating process

Our experimental environment is designed around a simple New Keynesian model that is commonly used for monetary policy analysis. We construct an economy that follows a data-generating process based on this canonical model (Woodford [2003]) that is calibrated to the Canadian economy, a model environment among many model candidates considered by the Bank of Canada for its mandate renewal during the 2021 exercise (Swarbrick and Zhang [forthcoming]). The economy in which participants interact is described by the following system of equations:

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa x_t + u_t \tag{1}$$

$$x_t = \mathbb{E}_t x_{t+1} - \frac{1}{\sigma}(i_t - \mathbb{E}_t \pi_{t+1} - r_t^n) \tag{2}$$

$$r_t^n = (1-\rho)(-ln(\beta)) + \rho r_{t-1}^n + \sigma_{rn}\epsilon_t \tag{3}$$

Equation 1 describes the evolution of inflation in period t, π_t, in response to aggregate one-period-ahead inflation expectations, $\mathbb{E}_t \pi_{t+1}$, and the output gap, deviations of output from its steady state level, x_t. The output gap, given by Equation 2, is a function of aggregate expectations of one-period-ahead inflation and output gap expectations, $\mathbb{E}_t x_{t+1}$, as well as the deviations of the nominal interest rate, i_t, from the natural rate of interest, r_t^n. The natural rate of interest, described by Equation 3, is the rate of interest that keeps the economy at full employment while keeping inflation constant. The natural rate of interest is assumed to follow an AR(1) process and is subject to a sequence of demand shocks, ϵ_t.[1] Parameters in our model are

[1]In our experiments, as in related New Keynesian learning-to-forecast experiments by Arifovic and Petersen [2017], Ahrens et al. [2017], Assenza et al. [2019], Hommes et al. [2019a], Kryvtsov and Petersen [2021]), we only focus on aggregate demand shocks. Pilot studies including supply shocks made the environment considerably complicated; we leave this to future research.

calibrated to quarterly data, as in Swarbrick and Zhang [forthcoming], and are consistent with Canadian data. The values of parameters are summarized in Table 1.

To close the model, we need to include a policy rule that governs the evolution of the nominal interest rate, i_t. The policy rule is our key source of experimental variation. We consider five distinct ad hoc policy rules. The first three mandates we consider involve the central bank targeting various metrics of inflation and the output gap.

Under IT and DM regimes, the central bank sets the nominal interest according to the following general policy rule:

$$i_t = \bar{r} + \phi_\pi(\pi_t - \pi^*) + \phi_x(x_t - x^*) \tag{4}$$

where it seeks to minimize deviations of inflation and output gap from their targeted level of zero. Parameters ϕ_π and ϕ_x govern the reactions of the central bank to deviations of inflation and output gap from their targeted levels. The difference between IT and DM is that the weight on the output gap, ϕ_x is assumed to be considerably larger and equal to ϕ_π under a dual mandate.

Under the AIT regime, the central bank sets the nominal interest rate to minimize deviations of inflation from its asymmetric inflation target based on the recent average inflation rate. The central bank also places some weight on the output gap when making its policy decisions. We consider two horizons for average inflation – short horizon of 4-quarter average and long horizon of 10-quarter average. Our objectives in studying two horizons in AIT is twofold. First, we would like to explore how AIT with different horizons performs in the lab and how this performance compares with the predictions of a model with rational expectations. Such results can be useful in guiding implementation of AIT by policy makers. Second, the theoretical prediction is that AIT approaches PLT when the horizon in computing average inflation goes to the limit of infinity. Therefore, AIT with longer horizon may be considered to be a more practical way to achieve results comparable to those in PLT without many practical challenges in implementing PLT (Amano et al. [2020]). Thus, our laboratory experiments can provide an indication of whether AIT could be used as a feasible practical alternative to implementation of PLT.

$$i_t = \bar{r} + \phi_\pi\left(\frac{\sum_{j=0}^{3} \pi_{t-j}}{4} - \pi^*\right) + \phi_x(x_t - x^*) \tag{5}$$

$$i_t = \bar{r} + \phi_\pi\left(\frac{\sum_{j=0}^{9} \pi_{t-j}}{10} - \pi^*\right) + \phi_x(x_t - x^*) \tag{6}$$

Under the two remaining policy mandates, the central bank instead aims to target pre-specified levels of price and nominal GDP targets. Under the price-level targeting mandate, the central bank responds to deviations of the price level, P_t from its targeted level, P^*, as well as the output gap:

$$r_t = \bar{r} + \phi_P(P_t - P^*) + \phi_x(x_t - x^*) \tag{7}$$

where $P_t = P_{t-1} + \pi_t$. Finally, a nominal GDP level targeting mandate involves the central bank instead adjusting nominal interest rates in response to deviations of the nominal GDP level, $NGDP_t$ from its targeted level, $NGDP^*$:

$$i_t = \bar{r} + \phi_{NGDP}(NGDP_t - NGDP^*) \quad (8)$$

where $NGDP_t = x_t + P_t$.

Parameters in the policy rules are derived from optimizing loss function as implemented by Swarbrick and Zhang [forthcoming]:

$$L = \sum_{t=1}^{50} \left(\pi_t^2 + x_t^2 + 0.5(i_t - i_{t-1})^2\right) \quad (9)$$

This ad hoc loss function gives a realistic description of the goals pursued by a central bank.The coefficients in the policy rules in different monetary policy regimes were chosen to optimize this loss function and are consistent with the theoretical horse race conducted in Swarbrick and Zhang [forthcoming].

3.2 Experimental implementation

Our experimental design follows closely the structure of previous New Keynesian learning-to-forecast experiments focused on expectation formation at the zero lower bound (Arifovic and Petersen [2017], Ahrens et al. [2017], Assenza et al. [2019], Hommes et al. [2019b], Kryvtsov and Petersen [2021]).

General game

Our experiment consists of six independent sessions for each of the six monetary policy treatments. For each session, we invited groups of seven inexperienced participants to play the roles of professional forecasters tasked with making forecasts in 50 sequential periods. In each period, each subject j submitted forecasts about inflation and the output gap in the subsequent period – $E_{j,t}\pi_{t+1}$ and $E_{j,t}x_{t+1}$. Actual outcomes for π_t, x_t, and i_t were determined based on the current period's realized ϵ_t and the median submitted forecasts for $t+1$ inflation and output gap according to 1-3 and one of the policy rules 4-8.

The exogenous shocks in the experimental economy were pre-drawn. This is described in the instructions to the participants. The shock sequence was chosen to implement two distinct phases in the experiment. Each session began with an initial stable phase during periods 1-19 and provided us with an opportunity to evaluate the relative performance of different policy mandates away from the effective lower bound. This phase was followed by a significant large negative demand shock in period 20 that brought the economy to the effective lower bound. The large negative demand shock dissipated rather quickly, returning to the steady-state level of zero by period 23. The second phase, which lasted the remaining 30 periods, enables us to study how economies respond to and recover under the different monetary policy mandates following episodes at the ELB. Thus, we can test the stabilization properties of the monetary policy regimes during stable and unstable periods, including periods at ELB. It is important to test the ability of different monetary policy regimes to handle ELB episodes, given that many economies are at ELB and many central banks conduct assessments of monetary policy toolkit.

Treatment implementation

All treatments were parameterized to have the same steady state. At the beginning of each session, participants observed a five-period path of all variables at their steady-state levels. Once the session began, participants observed new shocks to the natural rate of interest which would drive the economy out of the steady state.

Under rational expectations, past historical data would not matter when transitioning from the steady state to the first period of the game in all regimes except AIT. Historical data would not be used in the calculation of the policy rate for IT, DM, PLT, and NGDP. However, when calculating the policy rate in AIT-4 and AIT-10 regimes, average inflation is calculated over past realized inflation. In the simulations with rational expectations presented in Section 4, average inflation is computed using the assumption that inflation is equal to steady state in periods that have not yet realized within the experimental session ("non-realized"). For example, in the simulations of AIT-4 with rational expectations, in period $t = 1$, average inflation is computed as $(0 + 0 + 0 + \pi_t)/4$.[2] If we were to implement the AIT regimes in the same way as in the RE simulations, the potency of the response to actual inflation realized during the experiment would have been lower during initial periods of the experimental economy. The developments during the initial periods influence subjects' learning and the formation of their expectations. An insufficient policy response has the potential to create dynamics that were not conducive to subjects' understanding of the AIT regime and would risk the policy mandate's credibility.

Therefore, in implementing the AIT-4 and AIT-10 regimes, average inflation was computed using actual inflation realizations available during the experimental initial three or nine periods respectively.[3] This computation was explained in detail in the experimental instructions. In our view, this implementation gives the AIT policy rule the best chance to react to actual developments during the experiment to the best of its strength. Our approach to initial periods is preferable to using the assumption of steady-state inflation in "non-realized" periods.

Payoffs

Participants' earnings during the experiment are determined based on the accuracy of their forecasts, as in other learning-to-forecast experiments (e.g. Kryvtsov and Petersen [2021]. Subjects' payoff for period t was based on the accuracy of their forecasts made in period $t - 1$:

$$Payoff_{j,t} = 0.3 \left(2^{-.5|\mathbb{E}_{j,t-1}\{\pi_t\} - \pi_t|} + 2^{-.5|\mathbb{E}_{j,t-1}\{x_t\} - x_t|} \right) \tag{10}$$

Participants' total payoff over all the forecasting periods was converted to Canadian dollars at an exchange rate of 50 cents per point. Participants observe their total payoff during the experiment.

Instructions

Participants were provided with detailed information about the economy's data-generating process, including very clear descriptions of how the central bank would set monetary policy and its impact on the economy. This information was presented both descriptively and quantitatively in the form of explicit equations. We also explained how participants' forecasts would translate into points and payoffs at the end of the experi-

[2] In period $t = 2$, average inflation is computed as $(0 + 0 + \pi_{t=1} + \pi_{t=2})/4$, in period $t = 3$, average inflation is computed as $(0 + \pi_{t=1} + \pi_{t=2} + \pi_{t=3})/4$, and in period $t = 4$, average inflation is computed as $(\pi_{t=1} + \pi_{t=2} + \pi_{t=3} + \pi_{t=4})/4$. In all periods $t > 4$, average inflation is computed based on 4 periods.

[3] For example, in period $t = 1$, $\pi_{t=1}$ is used to compute deviation from inflation target, in period $t = 2$, average inflation over periods $t = 1$ and $t = 2$ is used $\frac{\pi_{t=1} + \pi_{t=2}}{2}$, etc. Starting from period $t = 4$, 4-period average inflation is used in AIT-4. And starting period $t = 10$,10-period average inflation is used in AIT-10.

ment. The experimental instructions can be found in the Online Appendix.

Information on the screen

Participants were presented with information about their experimental economy as it evolved during the experiment. Figure 1 shows the screenshot of the computer screen seen by the subjects during the experiment. They continuously observed 4 charts presenting shocks and interest rates, inflation, inflation target and the subject's own inflation forecast, output and the subject's own output forecast, nominal GDP level and price level, and the targets of the central bank (inflation in IT, DM, AIT-4, and AIT-10, as well as the price-level target in PLT and nominal output target in NGDP). The targets were displayed continuously as a horizontal line at zero (for inflation and output gap) and at 1000 for the price-level and nominal output targets.

On the left-hand side of the screen, there were two input windows where subjects submitted their forecasts of inflation and output. Subjects were given 75 seconds to submit their forecasts during the initial 10 periods and 50 seconds during the remaining periods of the experiment. If participants failed to input their forecast on time, the experiment would move on to the next round and they would simply earn zero points for their missed forecasts. The median forecast would instead be selected from the submitted forecasts. Subjects could submit any number they wished, positive, negative, or zero, with no upper or lower bounds on their forecasts.[4]

Experiments were conducted online with 252 undergraduate students from Simon Fraser University and Texas A&M University from May to July 2020 and in May–June 2021. The sessions for each treatment were equally split between the two institutions. Each session lasted approximately two hours. Subjects were paid a show-up fee of $10 in addition to pay linked to their performance, with an average total pay of $25.

4 Theoretical Predictions and Experimental Hypotheses

We formulate theoretical predictions about the stabilization performance of different monetary policy regimes based on the loss function in equation 9. We simulate our model with each monetary policy regime under rational expectations using the sequence of demand shocks implemented in the experiment. Then we compute average loss in each regime as a square root of loss (equation 9) divided by 50 periods; the results are presented in Table 2. The average total predicted losses over the 50 periods of shocks ranges from 23.7 to 28.1 basis points. We break down the total loss into the losses associated with deviations of inflation, output, and interest rates from the steady state (which in our case is the target) by policy mandate.

Given our simulated sequence of shocks, the overall total loss (as well as the loss associated with inflation) is predicted to be lowest under NGDP, followed closely by PLT. PLT is followed by AIT, with AIT with 10-period horizon performing better than AIT with 4-period horizon. DM and IT are predicted to produce relatively larger losses than other regimes. It should be noted that losses across these regimes are quite close in REE.

Owing to the nature of the various policy mandates, the policy regimes are predicted to generate noticeably

[4] Other papers in this literature commonly impose upper and lower bounds on forecasts, resulting in relatively more stable dynamics.

distinct aggregate dynamics. Figure 2 presents the rational expectation equilibrium solutions for inflation, output gap, and the nominal interest rate for considered monetary policy mandates associated with our pre-selected shock sequence. Note that while inflation deviates significantly more from the steady state under IT and DM than under NGDP and PLT, it is relatively more stable.

Using these simulations and calculated losses, we form our key testable hypothesis:

Hypothesis 1: The realized losses under the five mandates are ordered as follows $L_{NGDP} < L_{PLT} < L_{AIT-10} < L_{AIT-4} < L_{DM} < L_{IT}$.

4.1 Simulations with naive agents

Other experimental studies of monetary policy regimes illustrate that participants' expectations are mostly non-rational (Anufriev et al. [2013], Assenza et al. [2019], and others). Given this evidence, we introduce a very simple form of bounded rationality – naive expectations – into our model to understand the implications for stabilization properties of different monetary policy regimes. Naive expectations are set as $E_t\pi_{t+1} = \pi_{t-1}$ and $E_t x_{t+1} = x_{t-1}$. We find that the presence of naive agents can be disruptive to economies with certain monetary policy regimes. Level-targeting regimes such as PLT and NGDP can break down at certain shares of naive agents. The threshold share of naive agents is 33% in PLT regime and 45% in NGDP; economies become unstable with shares of naive agents above the threshold level. IT, DM, and AIT tolerate 100% of naive agents, remaining stable. In other words, PLT and NGDP are the least robust to the presence of naive expectations.

We have simulated our model with different shares of naive expectations using a sequence of demand shocks implemented in the experiment. Figure 3 presents losses from the simulations with rational expectations and simulations with different shares of naive agents – 33%, 45%, and 100%.[5]

The results presented in Figure 3 lead to the following interesting observations. First, the presence of naive agents leads to higher losses across all regimes. Second, the increase of the share of naive agents leads to the increase of the losses for all regimes (with the exception of AIT-10 following the shock that we will discuss below in section 4.1.1). Third, the ranking of monetary policy regimes changes with increase of the share of naive agents: performance of history-dependent regimes (AIT, PLT, NGDP) deteriorates relative to regimes responding to current inflation and output gap (IT and DM). We would like to note that AIT performs better than level-targeting regimes PLT and NGDP. These simple simulations with naive expectations illustrate the important role expectations play in the ability of different policy regimes to stabilize economy.

Next, we are going to discuss the mechanism through which naive expectations weaken the performance of history-dependent regimes. As described above, naive agents form their expectations for the next period based on the realization in the previous period. These expectations are purely backward looking and do not incorporate understanding of what a monetary policy regime aims to achieve (stabilize inflation and output) and how it works to achieve it. In other words, naive expectations do not have a forward-looking aspect and, as a result, they weaken the expectations channel on which history-dependent regimes rely for their superior performance in models with rational expectations. In addition, in the economy with naive agents, rational

[5]Tables summarizing losses in the simulations with 33%, 45%, and 100% of naive agents are in Appendix A.

agents account for non-rationality of naive agents and adjust their expectations relative to expectations in the model with only rational agents. Thus, the presence of naive agents diminishes the effectiveness of the expectations channel. Given that history-dependent regimes such as AIT, PLT, and NGDP rely heavily on the expectations channel, the performance of these regimes deteriorates as the expectations channel weakens.

The simulations with naive agents suggest that IT and DM may be more robust to the presence of non-rational expectations in their ability to stabilize the economy. And these simulations may be indicative of how these regimes may perform in the laboratory, where expectations are likely to deviate from rational.

4.1.1 Post-shock dynamics in the presence of naive agents

We would like to discuss two additional observations from Figure 3, both about performance of monetary policy regimes following the ELB shock. First, in the presence of naive agents (33%), NGDP performs better than other regimes, and notably better than PLT, another level-targeting regime. Second, performance of AIT-10 does not deteriorate monotonically with the increase in the share of naive agents in the post-shock period.

Post-shock dynamics in NGDP and PLT with naive agents

During a stable period, with the share of naive agents at 33%, NGDP performs somewhat better than PLT, and these two regimes outperform other regimes. However, following the ELB shock, performance of NGDP remains better than that of other regimes, while PLT performs much worse. In other words, during stable times, PLT can handle the presence of naive agents as well as NGDP does, but after ELB shock, PLT deteriorates substantially relative to NGDP and other regimes. Why can NGDP handle the period after ELB shock better than PLT? And why does PLT deteriorate so much? Deterioration of PLT's performance after ELB shock is mostly due to higher volatility of output (Table A1). The focus of PLT is stabilization of price level and making up for all past misses from the price-level target; therefore, price-level stability comes at a price of higher output volatility. In contrast, NGDP aims at stability of nominal output where deviations of price level can be compensated with deviations in output level, keeping nominal output stable. Indeed, NGDP has lower volatility of output than PLT during post-shock periods 20-50. And so, after ELB shock, PLT overreacts to deviations of price level from the target, leading to higher output volatility, thus reducing its stabilization performance. Such a focus on price level also reduces performance of PLT relative to IT and DM. It is worth noting that with an increase in the share of naive agents to the 45% threshold level (the threshold level for NGDP), performance of NGDP declines below that of all other regimes. Larger presence of naive agents further weakens the expectations channel and consequently, NGDP can no longer outperform other regimes.

Post-shock dynamics in AIT-10 with naive agents

The relationship between the presence of naive agents and the performance of AIT-10 is non-monotonic during the post-shock period: losses increase when the share rises from 0 to 33%, then decline when the share is 45% and increase when the share increases further to 100% (Figure 3, panel c), while during stable periods 1-19, an increase in the share of naive agents leads to the monotonic decline in performance of AIT-10 (Figure 3, panel b).

The presence of naive agents brings two effects. First, naive expectations carry over strength in inflation and output from past periods, which leads to smaller declines in these variables after ELB shock than in the case with rational expectations. Second, expectations channels weaken directly because of naive agents and indirectly because rational agents adjust their expectations to account for the presence of naive agents.

AIT-10 performs worse at 33% than at 0% because the presence of naive agents weakens the expectations channel (second effect dominates). When the share of naive agents increases from 33% to 45%, the first effect becomes more pronounced. As a higher share of backward-looking expectations carries over some of the past strength in economic variables, the ELB episode is less severe and shorter with AIT-10 than in other regimes, resulting in its better performance. However, when the share of naive agents reaches 100%, the complete absence of rational agents destroys forward-looking expectations and the expectations channel necessary for AIT-10 to work. As a result, at 100% AIT-10 performed worse than at 45%.

5 Results

5.1 Overview of aggregate dynamics

We begin the overview of our results with the presentation of the aggregate time series from the experimental economies. Figure 4 illustrates the dynamics of inflation, output, and interest rate in each of six sessions in treatments with DM and IT. The dynamics of inflation and output exhibit impressive consistency across six sessions for each of these regimes and are very similar to the simulations of our model in REE. Both DM and IT experience stable inflation and output and a brief episode at the ELB at the time of large demand shock. These economies recover relatively quickly from this shock, although somewhat more slowly than in the simulation with RE.

The stability of IT and DM in our experiments may be due to the relatively high responsiveness of policy to both output and inflation: this policy regime benefits from "divine coincidence" in the economy with only demand shocks. The policy rule in both IT and DM responds to both inflation and output, with relatively strong coefficients on both. The coefficients in our DM are the strongest considered in the literature. For example, Cornand and M'baye [2018] and Hommes et al. [2019b] study flexible IT mandate (interest rate responding to inflation and output gap) with a relatively small coefficient on output gap ($\phi_=0.5$), while Kryvtsov and Petersen [2013] consider coefficients of as high as $\phi_=1$. The simulations with naive agents discussed above suggest that IT and DM are more robust to the presence of non-rational expectations than other regimes, which may be the reason why these two regimes outperform others in the experiments with mostly non-rational expectations. Hommes et al. [2019b] show that in the presence of backward-looking expectations, a strong response to output is important for stabilizing output and inflation. We will discuss the role of backward-looking expectations in our experiments below.

Next, we turn our attention to the results from AIT-4 and AIT-10 treatments illustrated on Figure 5. The dynamics of inflation and output indicate that AIT-4 is capable of stabilizing economy similarly to IT and DM, whereas AIT-10 delivers less stability and reports less consistency across sessions than in IT, DM, and AIT-4. It is likely that there is more heterogeneity across subjects' forecasts within each session of

AIT-10 treatment than in IT, DM, and AIT-4 treatments. Our experimental results show that experimental economies perform relatively well in AIT treatments, especially in AIT with shorter horizon, AIT-4. This result is consistent with Amano et al. [2020], who find that in two-agent New Keynesian model with a fraction of backward-looking price setters, a shorter horizon is optimal in AIT.

Recent evidence from a household survey by Coibion et al. [2020a] indicates that people have difficulty understanding the AIT framework and do not react to the treatment with information about average inflation target differently than to information about inflation target. Our evidence from the experimental lab shows that AIT performs relatively well, suggesting that experimental participants understood AIT reasonably well, especially AIT of shorter horizon. Indeed, the outcomes in AIT-4 treatment are comparable to those in IT consistent with Coibion et al. [2020a], showing that people react similarly to AIT as they do to IT.

Finally, we present results from the level-targeting frameworks. Figure 6 presents results from the NGDP level-targeting treatment, and Figure 7 shows results from the PLT treatment. These figures present time series separately for the stable periods of experiment and for the remaining periods following ELB shock. During the stable periods, both NGDP and PLT show dynamics with larger amplitude of fluctuations than in the simulation with RE but with substantially more persistence than in the simulation. Before ELB shock, treatments with NGDP and PLT show more volatility and less consistency in dynamics than all other treatments. Following the ELB shock, all sessions in each of these regimes unravel into spiraling deflation and declining output. Such unraveling deflationary dynamics were not observed on other policy regimes in our experiments. Other experimental studies such as Arifovic and Petersen [2017], Hommes et al. [2019a], and Assenza et al. [2019] have reported evidence of deflationary spirals in the experimental economies that face ELB. Unraveling in these treatments may be a consequence of participants' panic as they observe their economies increasingly deviating from the central bank's announced targets.

5.2 The horse race

5.2.1 Macroeconomic stability of different regimes

Next, we summarize the performance of five monetary policy regimes in terms of their ability to stabilize inflation, output, and interest rate. We use the loss function (Equation 9) as a summary measure describing stabilization performance of each regime. Dynamics before and after shock are distinct; therefore, we assess performance for each sub-period. Figure 8 presents average loss for each monetary policy framework for pre-shock periods 1-19, and Figure 9 shows average losses during post-shock periods 20-50.[6]

Figure 8 and Figure 9 show very distinct rankings of policy regimes. This ranking is somewhat different before the ELB shock and after it. During the stable periods 1-19, AIT-4 and AIT-10 perform better than DM and IT, which are followed by NGDP and PLT. After the ELB shock, the performance of AIT-4 and AIT-10 deteriorates below that of DM and IT, which outperform NGDP and PLT. Overall, after the ELB shock, the rankings of the regimes decline as the degree of their history dependence increases. We perform a Wilcoxon rank order test (details are in Table 8) to assess whether the distribution of losses in each treatment has statistically identical medians. This table shows that losses in DM are statistically significantly different

[6]Y-axis is a logarithmic scale. Dots in the figures represent values outside upper adjacent value (upper quartile +3/2 Interquartile range).

from losses in AIT-10, AIT-4, NGDP, and PLT, and losses in AIT-4 and AIT-10 are statistically significantly different from losses in NGDP and PLT. Difference between losses in DM and IT, difference between losses in NGDP and PLT, and difference between losses in IT, AIT-4, and AIT-10 are not statistically significant.

The ranking of the regimes in the experiments is different from the ranking of the regimes in the model with REE (Table 2). A summary of the rankings of the regimes in the simulations with RE, naive expectations and in the experimental data is provided in Table 4. Based on the evidence from experiments, we reject our hypothesis about the relative stabilization performance of the six monetary policy frameworks outlined in Section 4. In the experiments, monetary regimes responding to concurrent inflation and output such as IT, DM, AIT-4, and AIT-10 outperform the most history-dependent regimes, PLT and NGDP. AIT-4 outperforms AIT-10, i.e. less history dependence results in better stabilization.

Overall, the performance of the regimes declines with increase in the extent of history dependence. Better performance of rate-targeting regimes IT, DM, AIT-4, and AIT-10 relative to level-targeting rules PLT and NGDP may be due to the divine coincidence in the model: in the presence of only demand shocks, a strong response of policy rule to both output and inflation achieves stabilization of both. In our parameterization, both IT and DM regimes benefit from responding relatively strongly to deviations to both inflation and output.

While our findings about performance of monetary policy regimes contrast with those in the model with RE, they are consistent with results from the horse race in ToTEM with estimated positive share of non-rational rule-of-thumb price and wage setters in the Canadian economy (Swarbrick and Zhang [forthcoming]).[7] Wagner et al. [forthcoming] confirm the ToTEM results regarding underperformance of highly history-dependent frameworks in a model with bounded rationality, or "cognitive discounting" (Gabaix [2020]). The presence of boundedly rational rule-of-thumb price setters weakens the performance of the history-dependent rule because it weakens the expectations channel. For example, in the PLT regime, a missing price-level target in the current period means that in the future, inflation must be higher to catch up with the target price level. Rational agents understand this mechanism and increase their inflation expectations, leading to higher inflation and output. However, the expectations channel does not work well with boundedly rational expectations. If inflation expectations are backward looking, agents do not raise inflation expectations in response to missed price-level targets or do not raise them sufficiently high. And thus the target continues to be missed without rational forward-looking expectations. This is especially important at ELB, where monetary policy cannot rely on lowering the interest rate to stimulate demand, and the expectations channel is key to managing real interest-rate expectations. Section 6.1.1 illustrates that experimental participants use forecasting rules that are mostly backward looking/trend chasing, i.e. not rational. Given that subjects in our experiments use non-rational expectations, our finding of better performance of DM and IT is consistent with Hommes et al. [2019b], who showed that "with non-rational expectations, inflation volatility can be lowered if the central bank reacts to output gap in addition to inflation."

[7] Rule-of-thumb price setters set prices either based on past inflation or inflation target not in the optimal way; for more details, see Dorich et al. [2013].

5.2.2 Is AIT good enough? Role of framing of policy regimes

We next discuss the role of horizon in monetary regimes and framing of regimes in their ability to stabilize economy. We have already shown that AIT of shorter horizon (AIT-4) performs overall better than AIT of longer horizon (AIT-10), consistent with findings in Amano et al. [2020] in the New Keynesian model in the presence of backward-looking agents and in experimental study by Salle [2021]. Theoretically, performance of AIT approaches that of PLT, as the horizon of AIT increases to infinite. But our experimental results show that both AIT-4 and AIT-10 perform better than PLT. Overall, performance declines with an increase in the extent of history dependence with the most history-dependent rule, PLT, performing the worst. This is likely because of the framing of these regimes: AIT is framed in terms of inflation rate, while PLT is framed in terms of price level. AIT is less cognitively demanding when participants are tasked with forecasting inflation. We will discuss reasons for weak performance of level-targeting regimes below. Here, we would like to note two conclusions. First, AIT is better than PLT if it is important to introduce history dependence. Second, longer history dependence in AIT is less stabilizing.

5.2.3 Price level and nominal output

Next, we will discuss dynamics of price level and nominal output in different monetary policy regimes. While only PLT targets price level, it is nevertheless interesting to understand the behavior of price-level across all six regimes and how close price level is to its target or level consistent with inflation target. Figure 10 presents average deviation of price level from price-level target in PLT and deviations of price level from price level consistent with inflation target in IT, DM, AIT-4, AIT-10, and NGDP. Pre-shock, level-targeting regimes PLT and NGDP are the most effective in stabilizing price level, as these regimes finish with the lowest deviations before ELB shock in period $t = 20$. Level-targeting regimes PLT and NGDP are followed by AIT-10 and AIT-4, as these regimes are closer in their design to PLT than IT and DM. Interestingly, during initial 10 periods, AIT-10 and AIT-4 are better in keeping price level close to the target than PLT. This happens as it likely takes a bit longer for learning to happen in PLT and for this regime come into its full force. And finally, IT and DM result in relatively higher deviations of price level than other regimes by the end of the stable period.

However, following the ELB shock, PLT loses its performance properties and becomes incapable of stabilizing price level, as this regime unravels into deflationary spirals. PLT does not perform well based on any metric, whether it is stabilization of price level or stabilization of inflation, output, and interest rate, as discussed above. NGDP performs even worse than PLT. IT, AIT-10, and AIT-4 are close in their ability to stabilize price level, as these regimes are more effective in keeping inflation stable. DM finishes below IT, AIT-10, and AIT-4, as this regime places lower weight on stabilizing inflation than they do.

Figure 11 illustrates deviations of nominal output from its target level in NGDP level targeting. This figure shows that NGDP can be effective in achieving nominal output target during the stable period, however, this performance is lost following the ELB shock: deviations of nominal output from target unravel as experimental economies with NGDP regime unravel into very strong deflationary spirals.

So far we have discussed macroeconomic stabilization properties of the six monetary policy regimes and how they vary during stable and unstable periods. One of notable observations is that level-targeting regimes do not perform as well as expected based on the theoretical model with rational expectations. Why does this happen?

6 Why do level-targeting regimes not work better?

The weak performance of history-dependent regimes in the experiments is due to a combination of participants having difficulty understanding these regimes ("don't get it") and the central bank having difficulty establishing their credibility ("don't buy it"). Limited comprehension of the regimes manifests itself in two ways: not enough participants forecast in the correct direction and, of those that do, forecasts fall short of what is rational and necessary to pull economies out of their deflationary spiral ("too little"). Even those who do try to forecast in the correction do so "too late."

The history-dependent regimes can be more difficult than IT or DM for participants to understand. The forecasting task in AIT, PLT, and NGDP regimes is more computationally and cognitively demanding than in IT and DM. In AIT, the participants need to keep track of average inflation over a certain number of periods and to figure out what inflation rate the policy rule aims for in the next period to make up for past misses. In PLT, participants need to track deviations from the target price level and figure out what deviations of the price level from the target imply for the inflation rate, and how the policy rule would achieve this target in future periods.

Similarly, in NGDP the participants need to interpret deviations of nominal output from its target to understand goals of policy rule for the next period's inflation and output. In all history-dependent regimes, the goal for the next period's inflation (and output) evolves over time, depending on the realized history. This is the key strength of such regimes in theoretical models with rational forward-looking expectations. However, history-dependent objectives can be a weakness working against these regimes in an experimental environment for several reasons. First, it is more challenging for participants to understand a time-varying policy goal. Second, time-varying goals do not possess the clarity and transparency of unchanging inflation targets as in IT and DM. The clearly communicated, constant target in IT and DM can serve as a salient focal point for anchoring expectations. Third, time-varying goals can erode the credibility of the regime by creating the impression of unfocused or wavering intentions. Lastly, these issues can be further exacerbated when the central bank fails to achieve these goals, with forecasters simply unable to "see it to believe it". Arifovic and Petersen [2017] find that the implementation of PLT by communicating time-varying targets for the inflation rate does not deliver the expected stability of this regime. Even when the subjects do not need to figure out the time-varying inflation target and are provided with this information by the experimenters, the PLT regime can be unstable.

6.1 Do people understand the monetary policy regimes?

The expectations channel is key to effective monetary policy and especially to the performance of history-dependent regimes. How do people set their expectations during the experiments? What are the differences in expectation formation across the competing regimes? Next, we present evidence about participants' expectations in the experimental data.

6.1.1 Forecasting heuristics

Our simulations in Section 4 suggest that the nature of expectation formation – rational versus naive – can crucially affect the relative performance of different monetary policy mandates. It is typically assumed that the way in which aggregate expectations are formed is policy invariant. However, Assenza et al. [2019] show that expectations of experimental participants tend to self-organize on different forecasting rules depending on monetary policy regime. We next consider how different monetary policy mandates influence participants' forecasting heuristics.

We use experimental data on participants' forecasts to determine how participants form their forecasts. We consider several types of expectation formation and assign a type to each participant that best fits their forecasting behaviour. Table 7 summarizes all mechanisms we have considered. As we discussed above, the experimental economies significantly deviate from rational expectations equilibrium paths. Therefore, we need to consider other types of forecasting mechanisms in addition to rational expectations. We study several heuristics that have been previously used in the literature on the formation of expectations in macroeconomic models. The simplest deviation from rational expectations is cognitive discounting à la Gabaix [2020], with expectations that are somewhat short (a share α) of a rational expectations solution. Cognitive discounting weakens the expectations channel and makes history-dependent rules less successful than rational expectations. We also consider a heuristic in which participants forecasts' are based on steady state/target.

We also consider backward-looking mechanisms in which the formation of expectations is history driven. Constant-gain learning has been widely used in the literature on the role of learning in macroeconomics (Evans and Honkapohja [2001]) and implications for monetary policy (Bullard and Mitra [2002]) and is supported by the evidence of such expectations in the survey data (Branch [2004]). We also consider trend-chasing expectations that have been shown to be used in the experimental data (Hommes et al. [2019b], Anufriev et al. [2013], Assenza et al. [2019] and others). Trend-chasing expectations nest naive expectations $E_{i,t}x_{t+1} = x_{t-1}$ under a trend-chasing parameter $\tau = 0$, and survey data provides empirical evidence of the use of naive expectations (Branch [2004]).

We determine the forecasting heuristic for each participant that best fits their forecasting behavior during each of the phases of the experiment. To do so, we compute the mean absolute error of each participant's expectations for each of the heuristics presented in Table 7. For some heuristics such as M2 Cognitive Discounting, M4 Constant Gain, and M5 Trend-Chasing, we consider a wide range of parameterizations. The cognitive discounting parameter α can take values in the range of $[0.1, 0.9]$. Constant gain parameter γ and trend chasing parameter τ are in the range of $[0, 1.5]$. We consider values of these parameters from these ranges with an increment of 0.1. We assign each participant the heuristic and its parameter value (if applicable) that produces the lowest mean absolute error. Participants' forecasts can be classified under different heuristics for two different phases of the experiment. The prevalence of the assigned heuristics, by treatment and phase, are presented in Figure 12 for inflation forecasts and Figure 13 for output forecasts.

The most striking result is the rarity of the rational expectations in the experimental data. Fewer than 5% of participants in any treatment can be classified as rational or *model consistent*, and in some cases, this share is close to zero. The consistent lack of rational expectations suggests that participants do not broadly appreciate how economic fundamentals influence aggregate dynamics. Participants' elicited expectations show little evidence that they made their forecasts in a forward-looking way responding to the dynamics of shocks and internalized the stabilizing role of monetary policy. There remains a possibility that participants may understand these elements but not fully, i.e. they may use cognitive discounting, which, like rational expectations, assumes agents are forward looking and respond solely to aggregate demand shocks, albeit in a more muted manner. However, we observe very little use of the cognitive discounting model. In most treatments, we observe under 5% of participants using cognitive discounting, and frequently this share is close to zero. We observe the highest incidence of cognitive discounting in the post-shock phase of the NGDP treatment, with 11% of participants. While this small minority of participants tried to use their understanding of this policy regime, their forecasts were insufficiently rational and their share too small to pull NGDP economies out of deflationary spiral in the post-shock phase.

We emphasize that although participants in the rate-targeting treatments are not especially model consistent in their forecasting, their beliefs are not wildly different from rational expectations. They are relatively well anchored on the inflation target. Moreover, the consistency in aggregate dynamics across sessions in rate-targeting treatments suggests a consistent aggregate forecasting heuristic, and not just random or confused expectations.

Backward-looking expectations – trend-chasing and constant gain learning – are the dominant forecasting heuristics in most of our treatments. Together, these backward-looking heuristics make up the majority of participants' forecasts. In inflation forecasts, the largest share of participants use trend chasing during both pre-shock and post-shock phases, with trend chasing becoming more prevalent post-shock with decline of constant gain learning. We observe similar composition and evolution of heuristics in output forecasts across all treatments, with the exception of AIT-4 and DM. AIT-4 and DM stand out among the treatments: the share of constant-gain learning is considerably larger than the share of trend-chasing in the pre-shock phase, although it declines below that of trend chasing post-shock.

Also noteworthy is the large minority of participants in the NGDP and PLT treatments classified as using steady-state/target forecasting during the post-shock phase. A detailed analysis of individual forecasts shows that only a small minority of these participants actually forecast the steady-state values of zero. Rather, they submit forecasts closer to the steady-state value of zero than to values implied by other heuristics, but their values are negative. This behavior is certainly not rational, as an ex-ante rational agent would anticipate very high levels of inflation and output gap given the observed deflation and negative output gaps.

How strong is trend chasing?

Most of the participants, irrespective of regime, form backward-looking, trend-chasing expectations. However, some of the regimes remain more stable than others, despite the marked absence of rational expectations in all of them. To better understand this observation, we explore how the degree of backward-looking expectations differs across the regimes. The most notable difference is in the strength of the trend-chasing parameter τ. We compare the empirical cumulative distribution function (CDF) of the trend-chasing parameter τ assigned to trend-chasing forecasters across treatments. Figure 14 plots these distributions for inflation and output forecasts by phase of experiment.

Pre-shock, there is relatively little difference across treatments in the distribution of the strength of the response to past trends. The median subject has an assigned τ parameter between 0.1 to 0.4, depending on the treatment. NGDP and PLT exhibit higher trend-chasing reactions compared with other treatments, but the differences aren't quantitatively large. The strength of trend chasing is relatively low in AIT-4, in addition to low incidence of trend-chasers in these regimes.

By contrast, in the post-shock phase, we observe considerable differences across treatments. Subjects in AIT-4, DM, IT, and even AIT-10 are substantially less responsive to past trends in inflation and output than level-targeting regimes PLT and NGDP. In IT, DM, AIT-4, and AIT-10, trend chasing is characterized by a median value of a trend-chasing parameter of roughly zero, indicative of simple naive forecasting. In PLT and NGDP, trend-chasing is very strong, with parameter τ close to or greater than 1. In the monetary policy regimes with more history dependence, participants react more to history following a large shock. This may indicate panic among participants after a large shock. It may also be difficult for participants to understand what these monetary policy regimes should do in such cases. Instead of correcting past misses by expecting reversal in inflation and output, participants use stronger trend chasing in an effort to catch up with economic dynamics. Strong trend-chasing forecasts further destabilize the economy and lead to explosive deflationary dynamics in PLT and NGDP regimes following ELB shock.

What are the implications of these strongly trend-chasing expectations for the stability of PLT? Honkapohja and Mitra [2014] suggest that PLT parameters of $\phi_P = 0.25$ and $\phi_x = 1$ would be sufficient to stabilize the economy if agents are rational or learn recursively (constant gain learning) to form expectations. By contrast, Hommes and Makarewicz [2021a] show that a non-linear New Keynesian system is stable when PLT responds very strongly to price deviations when agents are naive ($\phi_P = 3$ and $phi_x = 2$). Our experimental results suggest that forecasters are mostly trend chasing, with a relatively small share of naive expectations. The explosive dynamics in our experiments with PLT suggest that PLT might need to be even stronger than in the case of naive or constant-gain expectations in order to effectively manage expectations.

6.1.2 Deviations and dispersion of forecasts

We have just shown that most participants do not form rational expectations. How far are their expectations from rational model-consistent expectations? What is the extent of misunderstanding across forecasters? To assess the magnitude and scope of irrationality, we plot the median deviation from REE averaged across sessions for each treatment on Figure 15 and Figure 16. Summary statistics on average deviations and dispersion of forecasts for pre-shock and post-shock periods are presented in Table 8.

We observe substantial deviations from rational expectations and large dispersion among forecasters. Both pre-shock and post-shock, deviations from rationality are relatively lower in the rate-targeting regimes IT, DM, AIT-4, and AIT-10 (under 50 bps) compared with the level-targeting treatments PLT and NGDP (deviations ranging between 60 and 100 bps in the pre-shock period and exploding post-shock). Post-shock, forecasters in PLT and NGDP exhibit significant deviations from REE as both inflation and output become unstable.

A similar pattern exists for the dispersion of forecasts, with a much higher interquartile range among inflation and output forecasts in the level-targeting than those among rate-targeting regimes. Interestingly, during the post-shock period, the dispersion of forecasts declines in all rate-targeting regimes (except output forecast in AIT-10). Dispersion grows larger in PLT and NGDP post-shock as participants become uncertain about what values to forecast. On impact of the ELB shock, dispersion increased in all regimes (Figure 16). This finding is consistent with observations in surveyed expectations following the COVID-19 outbreak: consumer surveys have reported a sharp increase in divergence in consumers' inflation forecasts (CSC [2020]).

Under rational expectations, AIT performs closer to PLT as the length of the horizon increases. The evidence from our experiments suggests that AIT may be easier for people to understand than PLT. Both deviations from the rational solution and dispersion of forecasts are lower in the AIT treatments than in PLT, with lower deviations in the case of shorter horizon AIT-4 than longer horizon AIT-10. Pre-shock, forecasts in AIT-4 and AIT-10 are closer to rational than those in IT and DM. Post-shock, deviations in AIT-10 become larger than those in other rate-targeting regime, but deviations in AIT-4 remain smaller than DM although above those in IT. Taken together, AIT delivers macroeconomic stabilization better than PLT, as it is easier for subjects to understand this regime and, as a result, make better forecasts.

The evidence on forecasts deviations and their dispersion points to two key observations. First, it appears to be more difficult to understand PLT and NGDP than IT, DM, AIT-4, and AIT-10 as subjects' forecasts under level-targeting regimes are further from rational than under other regimes. Second, the misunderstanding of level-targeting regimes gives rise to greater heterogeneity among participants.

Bias in inflation and output forecasts
Table 8 shows that a positive bias in inflation expectations and negative bias in output gap expectations persists in most of our treatments. Inflation forecasts are above rational in all rate-targeting regimes IT, DM, AIT-4, and AIT-10, whereas output forecasts are below rational except in AIT-10. This suggests that participants tend to expect higher inflation and weaker output. This may be due to forecasters perceiving inflation as "bad for the economy," i.e. higher inflation is associated with weaker economic activity. Evidence of such views has been documented in survey data by Candia et al. [2020] among consumers across several countries, by Kamdar [2019] among consumers in the U.S., and by Coibion et al. [2020b] among firms in Italy.

The bias in the experiments may be a consequence of participants not fully appreciating the stabilizing effects of monetary policy on inflation. While participants are aware of how monetary policy directly affects output and how output contemporaneously affects inflation, the transmission of monetary policy to inflation may be less obvious. This limited ability of forecasters to connect the causal effects of monetary policy on inflation has also been documented in Mokhtarzadeh and Petersen [2020] and Kryvtsov and Petersen [2021].

6.1.3 Do people forecast in the direction of rational expectations?

While participants' expectations are substantially different from rational, to what extent do they form their forecasts in the direction of rational model-consistent expectations? We begin by considering a generous definition of *basic rationality* whereby participants forecast in the direction of the REE solution. We base this evaluation on the approach in Amano et al. [2011] who call it "directionality." Specifically, a forecast is considered in the direction of the REE solution if a participant forecasts higher (lower) values relative to the previous period's outcomes if the REE forecast is higher (lower) than the previous period's outcome.[8] The shares of inflation and output forecasts exhibiting basic rationality in our experiments are presented for each treatment and different phases in Table 10. The columns labelled *Both* report the share of both inflation and output forecasts made in the direction of the REE solution.

During the pre-shock phase, about 50–60% of inexperienced subjects exhibit basic rationality in forecasts of inflation or output, and about 30% of subjects demonstrate it for both forecasts. There is little difference in the prevalence of basic rationality across policy regimes. On impact of the shock, participants' basic rationality sharply increases in periods 20-21 when it comes to forecasting inflation, as participants observe an unexpected large negative demand shock. Interestingly, at the time of the shock, PLT and AIT-10 report the highest shares of basic rationality in inflation and output forecasts, NGDP has the highest basic rationality (85%) in inflation forecasts followed by PLT and DM, and AIT-10 has the highest presence of basic rationality in output forecasts (80%). Despite the sharp increase in the share of people forecasting in the correct direction in PLT and NGDP on the impact of the shock, these regimes start to destabilize following the shock. However, as the shock dissipates and fundamentals revert to the steady state, attention to the shock declines and the share of participants exhibiting basic rationality falls significantly in most treatments. The decline is most pronounced under the NGDP and PLT mandates for both inflation and output forecasts. Only 18% of NGDP participants and 26% of PLT participants forecast inflation and output in the correct direction in the post-shock phase. We attribute the relatively larger decline in basic rationality in these treatments to the higher cognitive load associated with processing these level-targeting regimes.

We have just seen that a relatively low share of participants forecast in the correct direction towards the REE solution. Another interesting question is about how close these forecasts are to REE. When participants forecast in the correct direction, do they go as far as REE? Or do their forecasts fall short? We next analyze the forecasting decisions of subjects exhibiting basic rationality and present their forecasts on Figures 18 and 19. These figures illustrate that forecasts of those forecasting in the correct direction of rational forecast fall short of what is required by REE: they are either too high, in the case of positive rational values, or too low, in the case of negative rational values. Forecasts of those without basic rationality deviate from rational expectations even more because they do not even move in the necessary direction. For example, in PLT, when the price level is above target, rational expectations suggest to expect lower inflation than in the previous period. Those with basic rationality forecast lower inflation than in the previous period, but their median forecast is not low enough to bring the price level back to target. And during the post-shock period, their forecasts do not incorporate fully the reversal prescribed by the PLT regime and therefore remain below rational value. Those without basic rationality do not even understand the reversal in inflation embedded in the PLT regime, and therefore, they do not lower their inflation forecasts below past-period inflation. As a result, forecasts of those with basic rationality are closer to rational solution than forecasts of those without basic rationality, but they fall short of what would be expected under rational forecasts.

Our analysis reveals two behavioural challenges for history-dependent mandates. First, in all treatments, the fraction of participants forecasting in the correct direction is relatively low. Second, even those who forecast in the correct direction do not forecast sufficiently close to the rational solution. These behavioural

[8] In Amano et al. [2011], subjects are defined as forecasting in the correct direction if in the regression analysis, subjects' forecasts are negatively related to past log deviation of price level from its target.

challenges result in macroeconomic dynamics that deviate from the rational solution and are less stable. The consequences of these two issues result in explosive dynamics in the level-targeting regimes PLT and NGDP. However, rate-targeting regimes appear more robust to lack of rationality.

Our results about basic rationality in IT and PLT are similar to those in Amano et al. [2011], who find that, on average, participants forecast in the correct direction in PLT. We would like to highlight the important difference between the experimental designs of Amano et al. [2011] and ours. In Amano et al. [2011], the experimental economy is exogenously generated and is simulated using a model with rational expectations, i.e. the expectations of their subjects' expectations do not feed into actual outcomes. In contrast, our experimental economy is self-referential: expectations feed into actual outcomes, and subjects observe actual outcomes and learn from them. In our experiments, as discussed, subjects' expectations are overwhelmingly non-rational, with stronger trend-chasing expectations in PLT (and NGDP) than in rate-targeting regimes. Non-rational expectations generate dynamics that are significantly different from dynamics in the model with rational expectations. For example, economies in PLT and NGDP treatments unravel into deflationary spirals. As a result, the forecasting task of subjects in our experiments could be considered more difficult than in Amano et al. [2011], and it may not be surprising that relatively small shares forecast in the correct direction. As dynamics in experimental economies deviate further from REE solutions, as in PLT, they lead participants' forecasts further from those consistent with rational expectations.

History-dependent policy mandates such as NGDP and PLT demand a high level of rationality to be effective at stabilizing the economy. Our results suggest there is an insufficient level of basic rationality, let alone full rationality, at both the individual and aggregate levels, to advocate enthusiastically for such level mandates.

6.1.4 Too little, too late: It does not work

Subjects react too little, too late in history-dependent regimes. On impact of the ELB shock in period $t = 20$, the share of basic rationality increases in inflation forecasts in PLT and NGDP and output forecasts in NGDP. Inflation forecasts become closer to REE (panel (a) of Figure 15), while output forecasts fall below REE forecast (panel (b) of 15) in all regimes. And while inflation and output forecasts recover towards REE in IT, DM, AIT-4, and AIT-10, they continue to unravel below REE in PLT and NGDP. Not enough people adjust their expectations in the direction of basic rationality. Quantitatively, expectations are not sufficiently adjusted upward towards REE to pull the economies out of the crisis. In other words, the expectations channel on which PLT and NGDP rely is *too little* or too weak. As a result, the ELB episode becomes much longer in PLT and NGDP than in other regimes. With time, as participants see their economy unravel into a deflationary spiral, basic rationality declines and deviations of forecasts from REE increase in PLT and NGDP.

6.1.5 Need to see it to believe it

Another challenge for history-dependent regimes is that participants often *"need to see it to believe it."* In the pre-shock phase, the deviations of inflation from the price level and NGDP level targets are large. By the time the large negative demand shock occurs, there is little credibility in the central bank's ability to achieve its target. As destabilization continues, people do not see that policy is working and, therefore, they do not believe it as evidenced by the decline of the share of participants exhibiting basic rationality when the deviations from target grow.

We next evaluate how a central bank's performance in achieving its targets influences its credibility. A candidate proxy for a subject's credibility in the central bank is if the subject forms her inflation expectations in line with the rational expectations solution. We use the indicator variable $\mathbb{1}_{i,t}^{BasicRationality}$ that takes the value of one if participant i exhibits basic rationality in period t when forming their $t+1$ forecast, as described in the previous subsections, and zero otherwise. Recent central bank performance, $AbsDevFromTarget_{t-1}$, is measured as the absolute deviation of the key macroeconomic variable from the central bank's stated target. For IT and DM, we calculate the absolute deviation of inflation from the inflation target of zero. For AIT-4 and AIT-10, we compare the average inflation over the past and current four and 10 periods to zero. For PLT and NGDP, we compare the price level and nominal GDP level to their respective targets of 1000. We also control for participant i's lagged absolute forecast errors $AbsFE\pi_{i,t-1}$, which we would expect to be negatively correlated with a participants' willingness to forecast in a model-consistent manner. Finally, we control for persistence in credibility by including a one-period lag of the indicator variable in our specifications. We estimate the following panel logit regressions by treatment over our pre- and post-shock data:

$$\mathbb{1}_{i,t}^{BasicRationality} = \alpha + \beta_1 \mathbb{1}_{i,t-1}^{BasicRationality} + \beta_2 AbsFE\pi_{i,t-1} + \beta_3 AbsDevFromTarget_{t-1} + \beta_4 \mu_i + \epsilon_{i,t} \quad (11)$$

where μ_i is the subject fixed effect and $\epsilon_{i,t}$ are robust standard errors. Table 10 presents the results by treatment and phase of the experiment.

Intriguingly, we find that credibility in IT and DM strengthens the further inflation is from its target. A one-basis-point increase in the deviation of inflation from target increases the log odds of forecasting rationally by 0.04 when participants are inexperienced. The effect is smaller but remains significant in the post-shock phase. Likewise, we observe a similar re-coordination on RE when the economy deviates further from the target in AIT-4 and AIT-10 in the pre-shock phase. These results suggest that larger, more salient deviations from the target are quickly corrected by participants adjusting their expectations in line with the central bank's target.

This is not the case in PLT and NGDP. Participants do not consistently respond to deviations from the target by adjusting their direction of forecasting. In fact, for inexperienced NGDP participants, credibility declines as the NDGP level deviates further from its target. Given the lack of credibility in the pre-shock phase, it is not surprising that credibility is weak after entering the ZLB and that monetary policy has no further capacity to provide stability.

We find that there is very strong persistence in subjects' basic rationality. If participants were previously forming expectations in line with rational expectations, and thus the central bank's mandate, they are more likely to continue to do so in most of our specifications. The persistence is strongest in IT, PLT, and NGDP, indicative of stable learning. By contrast, participants in AIT-4 and AIT-10 do not show significant persistence in their basic rationality and, in AIT-10, are more likely to switch away from rational forecasting if they had used it in the past.

Personal forecast errors do not play a consistent or large role in shaping participants' credibility in most cases. The exception is in the post-shock phase of IT, where participants who made larger forecast errors were more likely to form subsequent forecasts in the rational direction. Again, this is indicative of participants exhibiting a solid understanding of how monetary policy should influence inflation.

Both "don't get it" and "don't buy it" contribute to unstable dynamics in the PLT and NGDP treatments. Although participants form backward-looking, trend-chasing expectations in all regimes, PLT and NGDP give rise to the strongest extrapolation of past trends in the future. The trend-chasing expectations are more destabilizing in the level-targeting regimes than in other regimes.

7 Conclusion

Our assessment of five competing monetary policy frameworks suggests a distinct ranking based on their ability to achieve macroeconomic stabilization. Rate-targeting regimes such as IT, DM, and AIT outperform level-targeting regimes such as PLT and NGDP. The stabilization performance of the regimes declines with the horizon of history dependence after a large demand shock, bringing the economy to ELB.

Policymakers cannot take for granted that people form rational expectations. Our individual-level analysis of our experimental data suggests that people struggle with even the basic implications of monetary policy and do not internalize the goals of the competing policy regimes. Only a relatively small share of participants form expectations in the direction of the rational expectations solution ("basic rationality"). Forecasts of those with basic rationality quantitatively fall short of rational, i.e. they do not expect strong enough reversal in the dynamics. Furthermore, the share of the participants with basic rationality declines during the unstable post-shock period. In sum, these relatively rational forecasts are too few and not strong enough to reverse post-shock deflationary spirals.

Rather, the vast majority of our participants hold some form of trend-chasing or recursive-learning expectations. Moreover, there is a high dispersion in forecasts among participants, indicating heterogeneity in how they understand policy regimes and difficulty coordinating forecasts. While the composition of forecast heuristics is similar across all treatments during the initial stable periods of the experiments, it become distinct across treatments following the ELB shock. In particular, level-targeting regimes PLT and NGDP give rise to stronger trend chasing following the ELB shock. Forecast dispersion increases post-shock, and more so in level-targeting regimes. Instead of expecting corrections of past misses by monetary policy and reversals in inflation and output dynamics, participants use strong trend-chasing heuristics in an effort to catch up with evolving dynamics. This extrapolative behaviour further destabilizes economies into explosive dynamics with unraveling deflation in PLT and NGDP. Similar deflationary spirals have been observed in other experimental studies with an ELB: with IT and state-dependent inflation targeting in Arifovic and Petersen [2017] and IT in Hommes et al. [2019a] and Assenza et al. [2019].

Is AIT *good enough*? Our experimental results point to three conclusions regarding AIT. First, AIT is significantly better at stabilizing the economy than PLT, likely because the framing of policy in terms of a constant inflation target presents lower cognitive burden for the subjects. It may be easier for subjects to understand response to 4-quarter average inflation in AIT than the central bank's response to deviations of the price level from its target and what these deviations imply for next-period inflation rate in PLT. Second, a shorter history dependence in AIT improves macroeconomic stability. AIT-4 delivers more stable outcomes than AIT-10, which is consistent with the conclusion in Amano et al. [2020] that a short horizon is optimal for AIT in the presence of backward-looking agents and in an experimental study by Salle [2021]. Third, AIT with a short-horizon AIT-4 can stabilize the economy as well as IT or DM. The key underlying reason for these results is that non-rational backward-looking expectations are more disruptive for policy regimes with longer history dependence such as AIT-10 and PLT and NGDP.

Our experimental results may be useful for the practical implementation of AIT. Recent evidence from a household survey by Coibion et al. [2020a] indicates that U.S. consumers have difficulty understanding the AIT framework of the Federal Reserve, although evidence from Hoffmann et al. [2021] is more encouraging for AIT where German households revise their expectations consistently with AIT policy. Our evidence suggests that a longer-horizon AIT mandate may further complicate the public's understanding of the regime, which in turn can negatively affect its performance. In the case of the Federal Reserve, this task might be further complicated by the absence of a specific formula and horizon in the flexible framework of AIT introduced in August 2020 (Powell, Clarida [2021]). On the other hand, vague definition of policy goal may have some advantages in not losing credibility when goals are not achieved Stein [1989]. Orphanides and Williams [2005] illustrated that it is even more important to have credible inflation target under bounded rationality than under rational expectations.

We provide a broad experimental assessment of five general classes of monetary policy regimes. Our approach was to place these frameworks on as equal-footing as possible to understand their relative performance at managing expectations. We encourage further investigation into these frameworks, especially the ones that incorporate more history-dependence, to understand how different mandates should be implemented to achieve comparable results. This could be done through experimental variation in the complexity of the environment, parameterization of the mandates or shocks, participants' information sets and central bank communication.

In conclusion, our experimental results fail to provide support for the benefits of level-targeting frameworks such as PLT and NGDP targeting. These require much greater public education and communication strategies, possibly in the form of effective projections to guide expectations (Mokhtarzadeh and Petersen [2020], Kryvtsov and Petersen [2021]). The frontrunners of our horse race are IT, DM, and AIT with a short horizon. We consider these regimes closely related, as all of them are rate-targeting regimes and respond to both inflation and output. It may not be surprising, therefore, that these regimes perform similarly well in an environment with only demand shocks and deriving benefits from the "divine coincidence" of stabilizing both inflation and output. In future research, it would be interesting to extend this experimental framework to include supply shocks where NGDP has the potential to significantly dominate. However, based on the assessment in this study, switching from IT to level-targeting regimes does not appear desirable or advantageous.

References

Renewal of the Inflation-Control Target: Background Information — November 2011. *Bank of Canada*, 2011.

Canadian Survey of Consumer Expectations, Q2 2020. URL https://www.bankofcanada.ca/2020/07/canadian-survey-of-consumer-expectations-second-quarter-of-2020.

Klaus Adam. Experimental evidence on the persistence of output and inflation. *The Economic Journal*, 117 (520):603–636, 2007.

Steffen Ahrens, Joep Lustenhouwer, and Michele Tettamanzi. The stabilizing role of forward guidance: A macro experiment. 2017.

Robert Amano, Jim Engle-Warnick, and Malik Shukayev. Price-level targeting and inflation expectations: Experimental evidence. Technical report, Bank of Canada, 2011.

Robert Amano, Stefano Gnocchi, Silvain Leduc, and Joel Wagner. Average is good enough: Average-inflation targeting and the ELB. *Bank of Canada Staff Working Paper*, 2020-31, 2020.

Mikhail Anufriev, Tiziana Assenza, Cars Hommes, and Domenico Massaro. Interest rate rules and macroeconomic stability under heterogeneous expectations. *Macroeconomic Dynamics*, 17(8):1574–1604, 2013.

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

Tiziana Assenza, Peter Heemeijer, Cars H Hommes, and Domenico Massaro. Managing self-organization of expectations through monetary policy: A macro experiment. *Journal of Monetary Economics*, 2019.

Claes Berg and Lars Jonung. Pioneering price level targeting: The Swedish experience. *Journal of Monetary Economics*, 43:525–551, 1999.

Ben S. Bernanke. Monetary policy in a new era. *Paper prepared for conference on Rethinking Macroeconomic Policy, Peterson Institute, Washington, DC, October 12-13*, 2017.

William Branch. The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations. *Economic Journal*, 114(July):592–621, 2004.

James Bullard. A primer on price level targeting in the U.S. *Presentation, CFA Society of St. Louis, St. Louis, Mo.*, 2018. URL https://www.stlouisfed.org/from-the-president/speeches-and-presentations.

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

Bernardo Candia, Olivier Coibion, and Yuriy Gorodnichenko. Communication and the beliefs of economic agents. *NBER Working Paper*, 2020.

Mark Carney. Guidance. *Speech, CFA Society Toronto, December 12*, 2012.

Richard H. Clarida. The Federal Reserve's new framework: Context and consequences. *At "The Road Ahead for Central Banks," a seminar sponsored by the Hoover Economic Policy Working Group, Hoover Institution, Stanford, California (via webcast)*, 2021. URL https://www.ecb.europa.eu/press/key/date/2020/html/ecb.sp200930~169abb1202.en.htmlhttps://www.federalreserve.gov/newsevents/speech/clarida20210113a.htm.

Olivier Coibion, Yuriy Gorodnichenko, Edward S. II Knotek, and Raphael Schoenle. Average inflation targeting and household expectations. *Manuscript*, 2020a. URL https://eml.berkeley.edu/~ygorodni/CGKS_AIT.pdf.

Olivier Coibion, Yuriy Gorodnichenko, and Tiziano Ropele. Inflation expectations and firms' decisions: Causal evidence. *Quarterly Journal of Economics*, 135:165–219, 2020b.

Camille Cornand and Cheick Kader M'baye. Does inflation targeting matter? An experimental investigation. *Macroeconomic Dynamics*, 22(2):362–401, 2018.

Jose Dorich, Michael Johnston, Rhys Mendes, Stephen Murchison, and Yang Zhang. ToTEM II: An Updated Version of the Bank of Canada's Quarterly Projection Model. *Bank of Canada Technical Report No. 100*, October 2013. https://www.bankofcanada.ca/wp-content/uploads/2013/10/technical_report_100.pdf.

John Duffy. Macroeconomics: A survey of laboratory research. *Handbook of Experimental Economics*, 2: 1–90, 2016.

Charles M. Evans. Monetary policy in low-inflation environment: Developing a state-contingent price-level target. *Journal of Money, Credit and Banking*, 44(1):147–155, 2012.

George Evans and Seppo Honkapohja. *Learning and Expectations in Macroeconomics*. Princeton University Press, 2001.

George W. Evans, Seppo Honkapohja, and Ramon Marimon. Convergence in monetary inflation models with heterogeneous learning rules. *Macroeconomic Dynamics*, 5(1):1–31, 2001.

Xavier Gabaix. A behavioral New Keynesian model. *American Economic Review*, 110(8):2271–2327, 2020.

Mathias Hoffmann, Emanuel Moench, Lora Pavlova, and Guido Schultefrankenfeld. Would households understand average inflation targeting? *Manuscript*, 2021.

Cars Hommes and Tomasz Makarewicz. Price level versus inflation targeting under heterogeneous expectations: A laboratory experiment. *Journal of Economic Behavior and Organization*, 182:39–82, 2021a. ISSN 0167-2681. doi: https://doi.org/10.1016/j.jebo.2020.11.011. URL http://www.sciencedirect.com/science/article/pii/S0167268120304133.

Cars Hommes and Tomasz Makarewicz. Experimental study on the effect of nominal price level versus inflation targeting with and without guidance. *Journal of Economic Behavior and Organization*, 2021b.

Cars Hommes, Domenico Massaro, and Isabelle Salle. Monetary and fiscal policy design at the zero lower bound: Evidence from the lab. *Economic Inquiry*, 57(2):1120–1140, 2019a.

Cars Hommes, Domenico Massaro, and Matthias Weber. Monetary policy under behavioral expectations: Theory and experiment. *European Economic Review*, 118:193–212, 2019b.

Seppo Honkapohja and Kaushik Mitra. Targeting nominal GDP or prices: Guidance and expectation dynamics. *Bank of Finland Research Discussion Papers*, 2014.

Seppo Honkapohja and Kaushik Mitra. Price level targeting with evolving credibility. *Journal of Monetary Economics*, 116:88–103, 2020.

Lars Jonung. Knut Wicksell's norm of price stabilization and Swedish monetary policy in the 1930's. *Journal of Monetary Economics*, 5:459–496, 1979.

Rupal Kamdar. The inattentive consumer: Sentiment and expectations. *Manuscript*, 2019. URL https://rupalkamdar.github.io/pdfs/Inattentive_Consumer.pdf.

Oleksiy Kryvtsov and Luba Petersen. Expectations and monetary policy: Experimental evidence. Technical report, Bank of Canada, 2013.

Oleksiy Kryvtsov and Luba Petersen. Central bank communication that works: Lessons from lab experiments. *Journal of Monetary Economics*, 117:760–780, 2021.

Kaushik Mitra. Desirability of nominal GDP targeting under adaptive learning. *Journal of Money, Credit and Banking*, 2003.

Fatemeh Mokhtarzadeh and Luba Petersen. Coordinating expectations through central bank projections. *Experimental Economics*, 24(3):1–36, 2020.

Athanasios Orphanides and John Williams. Imperfect knowledge, inflation expectations and monetary policy. In Ben S. Bernanke and Michael Woodford, editors, *Inflation Targeting*. University of Chicago Press, 2005.

Damjan Pfajfar and Blaž Žakelj. Experimental evidence on inflation expectation formation. *Journal of Economic Dynamics and Control*, 44:147–168, 2014.

Jerome H. Powell. New economic challenges and the Fed's monetary policy review. *Speech at "Navigating the Decade Ahead: Implications for Monetary Policy," an economic policy symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, August 27, 2020*. URL https://www.federalreserve.gov/newsevents/speech/powell20200827a.htm.

Isabelle Salle. What to target? Insights from theory and lab experiments. *Bank of Canada Staff Working Paper*, 2021.

Thomas Sargent. *Bounded Rationality in Macroeconomics*. Clarendon Press, 1993.

Thomas Sargent. *The Conquest of American Inflation*. Princeton University Press, 1999.

Jeremy Stein. Cheap talk and the Fed. *American Economic Review*, 79(1):32–42, 1989.

Jonathan Swarbrick and Yang Zhang. A horse race of monetary policy strategies for Canada. *Bank of Canada Staff Working Paper*, forthcoming.

Joel Wagner, Yang Zhang, and Thomas Schlanger. A horse race of alternative monetary policy regimes under bounded rationality. *Bank of Canada Staff Working Paper*, forthcoming.

John C. Williams. Preparing for the next storm: Reassessing frameworks and strategies in a low r-star world. *FRB Economic Letter*, 2017-13:1–8, 2017.

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

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

Table 1: Model and experiment parameters

	Parameter	Value
All regimes	β	0.994
	σ	1
	ρ	0.8
	σ_{rn}	0.005
	κ	0.125
	π^*	0
	x^*	0
	r_t^{n*}	0
	$\bar{r} = i^*$	60
IT	ϕ_π	5.5
	ϕ_x	3.0
AIT	ϕ_π	5.5
	ϕ_x	3.0
DM	ϕ_π	4.5
	ϕ_x	4.5
PLT	ϕ_P	0.8
	ϕ_x	1.3
NGDP level	ϕ_{NGDP}	1.1

Parameterizations used for the general data-generating process and for the individual policy mandate treatments.

Table 2: Losses associated with inflation, output, and interest rate in REE

	$\sqrt{\text{total loss}/T}$	$\sqrt{\sum^T (\pi - \pi^*)^2/T}$	$\sqrt{\sum^T (x - x^*)^2/T}$	$\sqrt{\sum^T (i - i^*)^2/T}$
periods 1-50				
NGDP	168.2	14.8	82.8	206
PLT	169.8	9.3	40.4	232.8
AIT-10	179.7	20.2	56.5	239.6
AIT-4	180.8	22	47.4	244.8
DM	184.4	21.7	38.1	253.4
IT	186.9	24.8	43	254.8
periods 1-19				
NGDP	153.8	11.8	75.5	188.8
PLT	155.9	7.5	34.3	214.8
AIT-10	164	18.2	49.8	219.5
AIT-4	165.3	20	42.1	224.2
DM	168.5	19.6	32.1	232.4
IT	170.8	22.4	36.8	233.7
periods 20-50				
NGDP	176.4	16.4	87	215.8
PLT	177.7	10.2	43.7	243.2
AIT-10	188.7	21.3	60.2	251.2
AIT-4	189.7	23.1	50.3	256.6
DM	193.5	22.8	41.3	265.4
IT	196.2	26.1	46.5	267

Loss and standard deviations of inflation, output, and interest rate were computed from simulations with rational expectations and are expressed in basis points.

Table 3: Losses in the laboratory experiments

regime	periods 1-50	periods 1-19	periods 20-50
AIT-4	172	154	182
DM	176	168	181
IT	177	170	181
AIT-10	186	152	203
PLT	$31*10^9$	213	$39*10^9$
NGDP	$4*10^1 5$	221	$5*10^1 5$

This table presents losses averaged across all sessions of each treatment.

Table 4: Ranking of regimes in the simulations with RE, naive expectations and in the data from laboratory experiments

ranking	REE	naive =33%	lab data
	Table 2	Table A1	Table 3
Periods 1-19			
1	NGDP	NGDP	AIT-10
2	PLT	PLT	AIT-4
3	AIT-10	AIT-10	DM
4	AIT-4	IT	IT
5	DM	DM	PLT
6	IT	AIT-4	NGDP
Periods 20-50			
1	NGDP	NGDP	IT
2	PLT	IT	DM
3	AIT-10	DM	AIT-4
4	AIT-4	AIT-4	AIT-10
5	DM	PLT	PLT
6	IT	AIT-10	NGDP
All Periods			
1	NGDP	NGDP	AIT-4
2	PLT	IT	DM
3	AIT-10	DM	IT
4	AIT-4	AIT-4	AIT-10
5	DM	PLT	PLT
6	IT	AIT-10	NGDP

Note: Ranking in column "naive=33%" is based on simulations with 33% of naive agents in all regimes. For details on losses for these simulations, see Table A1 in the Appendix.

Table 5: Wilcoxon rank order test, statistical significance

Periods 1-19						
	NGDP	PLT	DM	IT	AIT-4	AIT-10
NGDP						
PLT	0.87					
DM	0.0163	0.037				
IT	0.0782	0.0542	0.8728			
AIT-4	0.0104	0.0103	0.1093	0.3367		
AIT-10	0.0039	0.0161	0.0547	0.4233	0.5218	
Periods 20-50						
	NGDP	PLT	DM	IT	AIT-4	AIT-10
NGDP						
PLT	0.4225					
DM	0.0039	0.0039				
IT	0.0039	0.0039	0.631			
AIT, short	0.0039	0.0039	0.7488	0.631		
AIT-4	0.0039	0.0039	0.0065	0.0065	0.0104	
AIT-10						
	NGDP	PLT	DM	IT	AIT-4	AIT-10
NGDP						
PLT	0.4225					
DM	0.0039	0.0039				
IT	0.0039	0.0039	0.8728			
AIT-4	0.0039	0.0039	0.3367	0.3367		
AIT-10	0.0039	0.0039	0.2002	0.2002	0.025	

Results from Wilcoxon test based on the average losses from each of 6 sessions for all treatments. These results are for the hypothesis that losses in the treatments listed in the rows are equal to the losses in the treatments listed in the columns.

Table 6: Wilcoxon rank order test, probability that regimes in rows have lower losses than regimes in columns

	NGDP	PLT	DM	IT	AIT-4	AIT-10
Periods 1-19						
	NGDP	PLT	DM	IT	AIT-4	AIT-10
NGDP						
PLT	0.528					
DM	0.917	0.861				
IT	0.806	0.833	0.472			
AIT-4	0.944	0.944	0.778	0.667		
AIT-10	1	0.917	0.833	0.639	0.389	
Periods 20-50						
	NGDP	PLT	DM	IT	AIT-4	AIT-10
NGDP						
PLT	0.639					
DM	1	1				
IT	1	1	0.417			
AIT-4	1	1	0.444	0.417		
AIT-10	1	1	0.028	0.028	0.056	
All Periods						
	NGDP	PLT	DM	IT	AIT-4	AIT-10
NGDP						
PLT	0.639					
DM	1	1				
IT	1	1	0.472			
AIT-4	1	1	0.667	0.667		
AIT-10	1	1	0.278	0.278	0.111	

Results from Wilcoxon test based on the average losses from each of 6 sessions for all treatments. These results present the probability that losses in the treatments listed in the rows are less than the losses in the treatments listed in the columns, in accordance with hypothesis H1 presented in Section 4.

Table 7: Forecasting heuristics

Model	Heuristic Name	Model
M1	Ex–Ante Rational	$E_{i,t}x_{t+1} = f(r^n_{t-1}, \epsilon_t)$ $E_{i,t}\pi_{t+1} = f(r^n_{t-1}, \epsilon_t)$
M2	Cognitive Discounting	$E_{i,t}x_{t+1} = \alpha f(r^n_{t-1}, \epsilon_t)$ $E_{i,t}\pi_{t+1} = \alpha f(r^n_{t-1}, \epsilon_t)$
M3	Constant Gain	$E_{i,t}x_{t+1} = E_{i,t-1}x_t - \gamma(E_{i,t-2}x_{t-1} - x_{t-1})$ $E_{i,t}\pi_{t+1} = E_{i,t-1}\pi_t - \gamma(E_{i,t-2}\pi_{t-1} - \pi_{t-1})$
M4	Steady State/Target	$E_{i,t}x_{t+1} = 0$ $E_{i,t}\pi_{t+1} = 0$
M5	Trend Chasing	$E_{i,t}x_{t+1} = x_{t-1} + \tau(x_{t-1} - x_{t-2})$ $E_{i,t}\pi_{t+1} = \pi_{t-1} + \tau(\pi_{t-1} - \pi_{t-2})$

Models of expectations as functions of exogenous or historical data.
$\alpha \in [0.1, 0.9]$, γ and $\tau \in [0, 1.5]$ in increments of 0.1.

Table 8: Summary statistics about forecasts

	Deviation from REE		Interquartile Range	
	$E_{i,t}\pi_{t+1}$	$E_{i,t}x_{t+1}$	$E_{i,t}\pi_{t+1}$	$E_{i,t}x_{t+1}$
Periods 1-19				
AIT-4	29.02	-4.51	17.94	27.26
	(13.92)	(16.5t)	(30.57)	(37.76)
AIT-10	37.81	13.06	16.35	18.73
	(16.89)	(28.22)	(22.99)	(32.52)
DM	46.79	-26.53	12.15	18.76
	(17.67)	(17.33)	(18.05)	(36.20)
IT	37.11	-34.48	22.93	20.93
	(31.70)	(38.28)	(37.11)	(22.32)
NGDP	51.21	-153.5	33.97	59.54
	(84.32)	(190.2)	(28.91)	(38.90)
PLT	53.19	-16.07	49.98	63.06
	(66.38)	(87.53)	(49.4)	(51.08)
Periods 20-50				
DM	31.56	-20.21	8.483	9.906
	(13.17)	(17.91)	(17.58)	(12.65)
IT	14.89	-27.7	11.66	17.45
	(19.69)	(41.03)	(11.64)	(19.52)
AIT-4	25.71	-8.463	10.77	22.77
	(15.63)	(31.41)	(9.791)	(22.01)
AIT-10	44.55	6.18	14.74	22.6
	(49.99)	(43.14)	(10.49)	(17.28)
NGDP	-5.4e+14	-8.9e+14	1.24E+25	1.24E+25
	(5.3e+15)	(7.6e+15)	(2.14E+26)	(2.14E+26)
PLT	-1.6e+09	-3.4e+09	1.79E+09	2.75E+09
	(1.4e+10)	(3.2e+10)	(2.18E+10)	(2.23E+10)

This table presents inflation and output gap forecast statistics. Columns (1) and (2) present the average across sessions of the median absolute forecast deviations from the REE solution. Columns (3) and (4) present the interquartile range of forecasts. Standard deviations are presented in parentheses.

Table 9: Share of forecasts exhibiting basic rationality

	Preshock (Periods 1-19)			Shock (Periods 20-21)			Postshock (Periods 22-50)		
	Inflation	Output	Both	Inflation	Output	Both	Inflation	Output	Both
NGDP	0.59	0.49	0.26	0.85	0.22	0.12	0.29	0.29	0.18
PLT	0.49	0.64	0.29	0.71	0.58	0.49	0.36	0.38	0.26
DM	0.47	0.49	0.25	0.63	0.56	0.27	0.36	0.57	0.16
IT	0.54	0.54	0.33	0.70	0.49	0.26	0.46	0.54	0.21
AIT-4	0.48	0.62	0.32	0.58	0.52	0.20	0.43	0.56	0.25
AIT-10	0.48	0.57	0.33	0.63	0.80	0.51	0.51	0.65	0.36

This table presents the share of forecasts in the direction of the rational expectations equilibrium solution. A forecast exhibits *basic rationality* if it is higher (lower) than the previous realized outcome when the predicted REE is above (below) the previous outcome. *Both* indicates the share of simultaneously submitted inflation and output forecasts both exhibiting basic rationality.

Table 10: Central bank credibility - logit results with additional controls

Dep. Var.	Preshock: Periods 1-19						Postshock: Periods 20-50					
$\mathbb{1}_{i,t}^{BasicRationality}$	IT	DM	AIT-4	AIT-10	PLT	NGDP	IT	DM	AIT-4	AIT-10	PLT	NGDP
$\mathbb{1}_{i,t-1}^{BasicRationality}$	0.920***	0.609***	0.043	-0.510*	0.710***	0.657***	0.539***	0.772***	0.463***	0.839***	1.579***	2.533***
	(0.17)	(0.17)	(0.22)	(0.26)	(0.17)	(0.17)	(0.12)	(0.12)	(0.13)	(0.13)	(0.13)	(0.20)
$\lvert E_{i,t-2}\pi_{t-1} - \pi_{t-1}\rvert$	-0.000	-0.007	-0.001	0.001	0.001	-0.004	0.023***	-0.002	0.000	-0.005	0.000	
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	
$\lvert \pi_{t-1} - \pi^*\rvert$	0.040***	0.047***					0.015***	0.020***				
	(0.01)	(0.01)					(0.00)	(0.00)				
$\lvert \frac{1}{4}\sum_{t-4}^{t-1}\pi_j - \pi^*\rvert$			0.088***						0.006			
			(0.02)						(0.00)			
$\lvert \frac{1}{10}\sum_{t-10}^{t-1}\pi_j - \pi^*\rvert$				0.051***						0.001		
				(0.01)						(0.00)		
$\lvert P_{t-1} - P^*\rvert$					-0.001						0.000	
					(0.00)						(0.00)	
$\lvert NGDP_{t-1} - NGDP^*\rvert$						-0.001*						-0.000
						(0.00)						(0.00)
α												-1.779***
												(0.14)
N	703	645	529	339	661	664	1278	1278	1167	1185	1209	1035
χ^2	87.96	32.49	34.80	29.53	20.96	25.97	47.68	66.06	15.68	48.26	147.5	.
p	6.01e-19	4.13e-7	1.34e-7	1.73e-6	0.000107	9.68e-6	2.49e-10	2.97e-14	0.00132	1.87e-10	6.15e-34	.

This table presents results from a series of fixed-effects probit panel regressions. The dependent variable is an indicator variable that takes the value of 1 if participant i in period t inflation exhibits basic rationality. α denotes the estimated constant. NGDP post-shock results are estimated and reported with a random effects specification due to convergence issues. *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Figure 1: Screenshot of participants' screens during the experiment

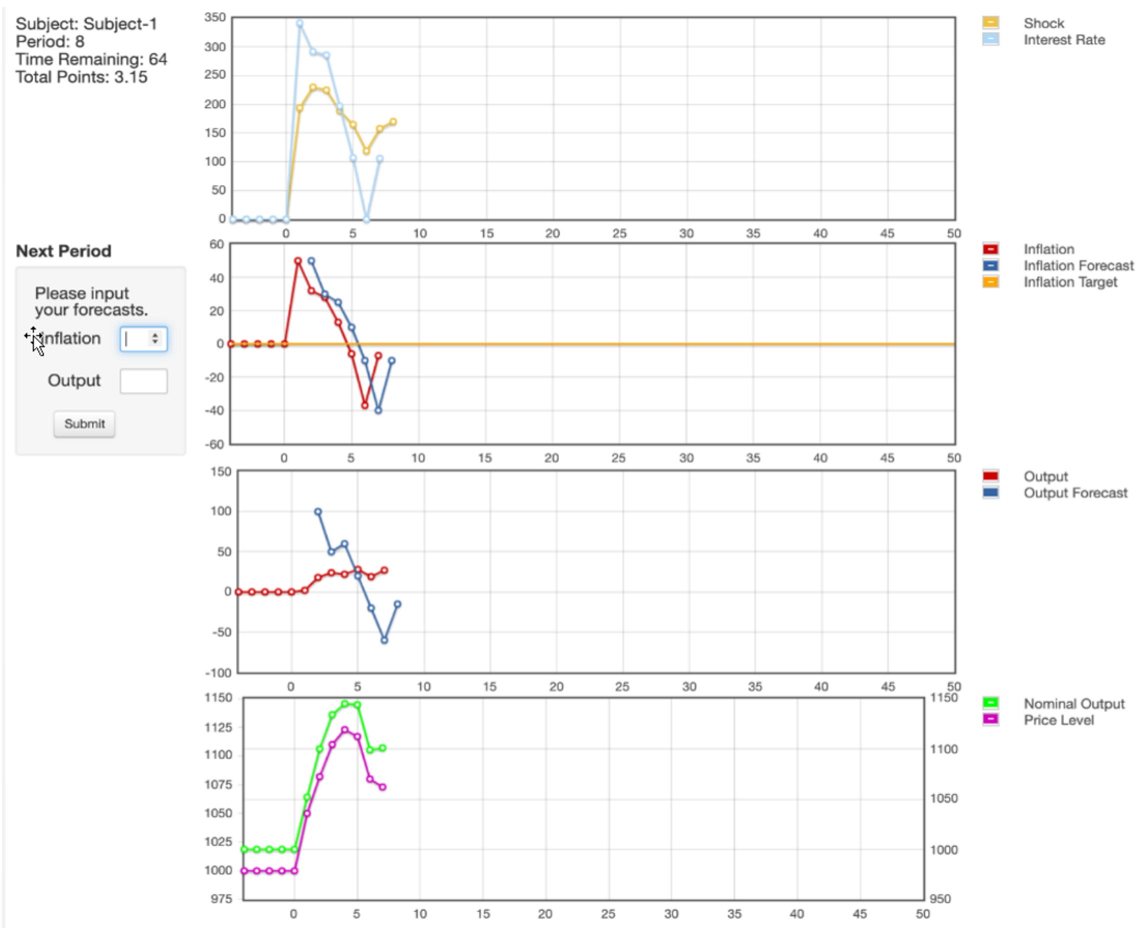


Figure 2: Simulations with rational expectations, New Keynesian model with ELB, and shock sequence used in the lab

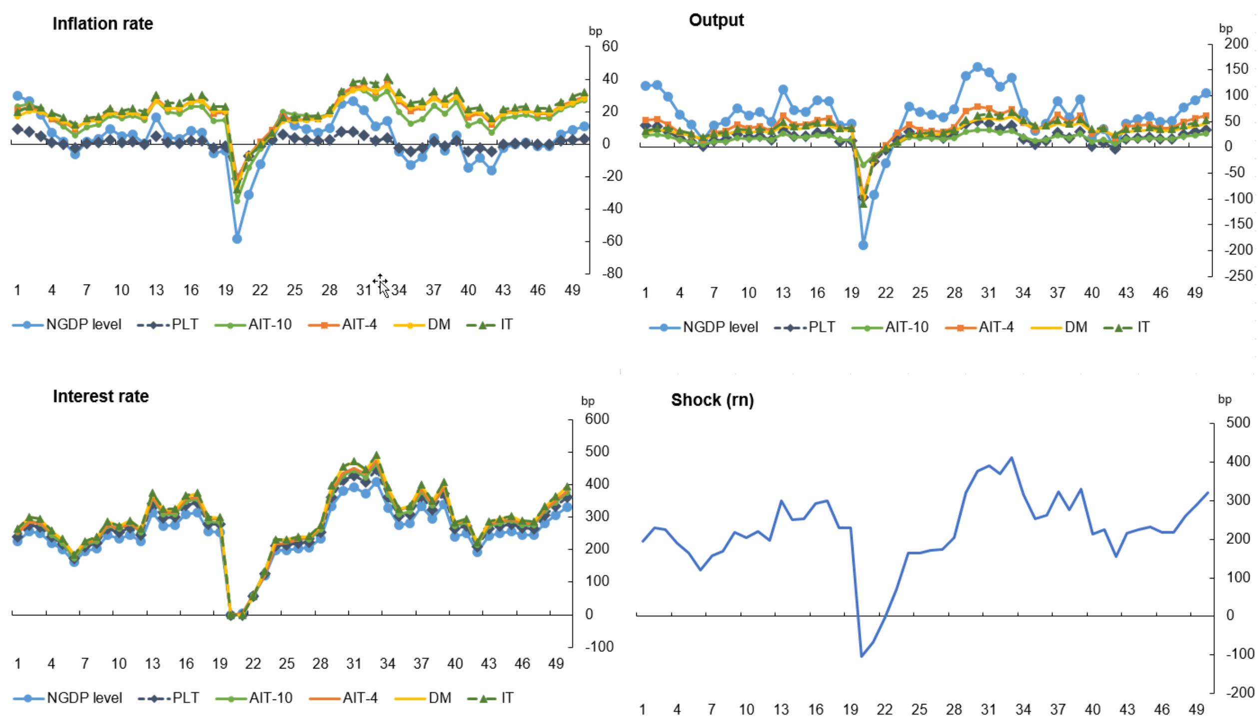


Figure 3: Summary of losses from simulations with rational expectations and simulation with naive agents

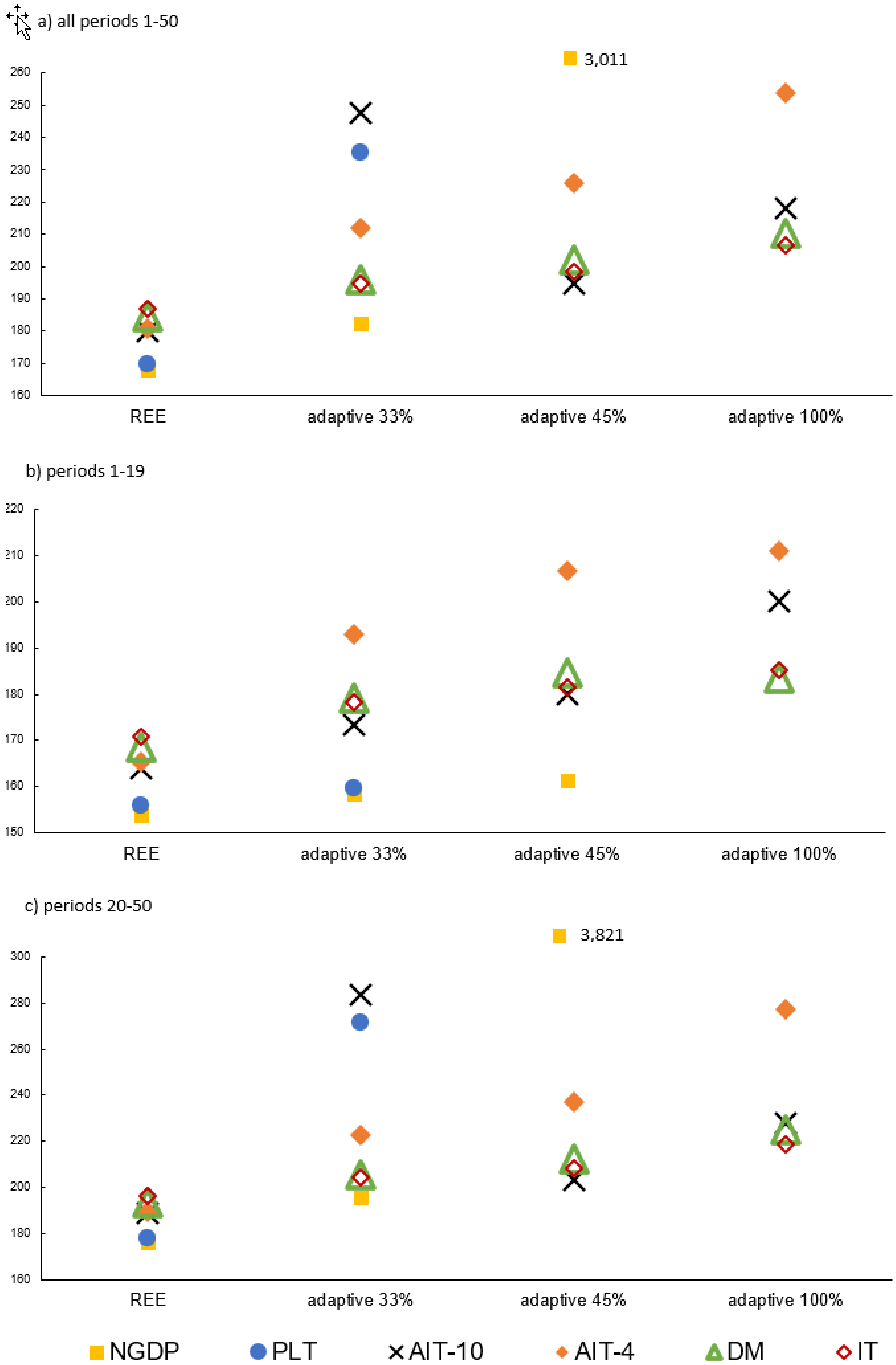
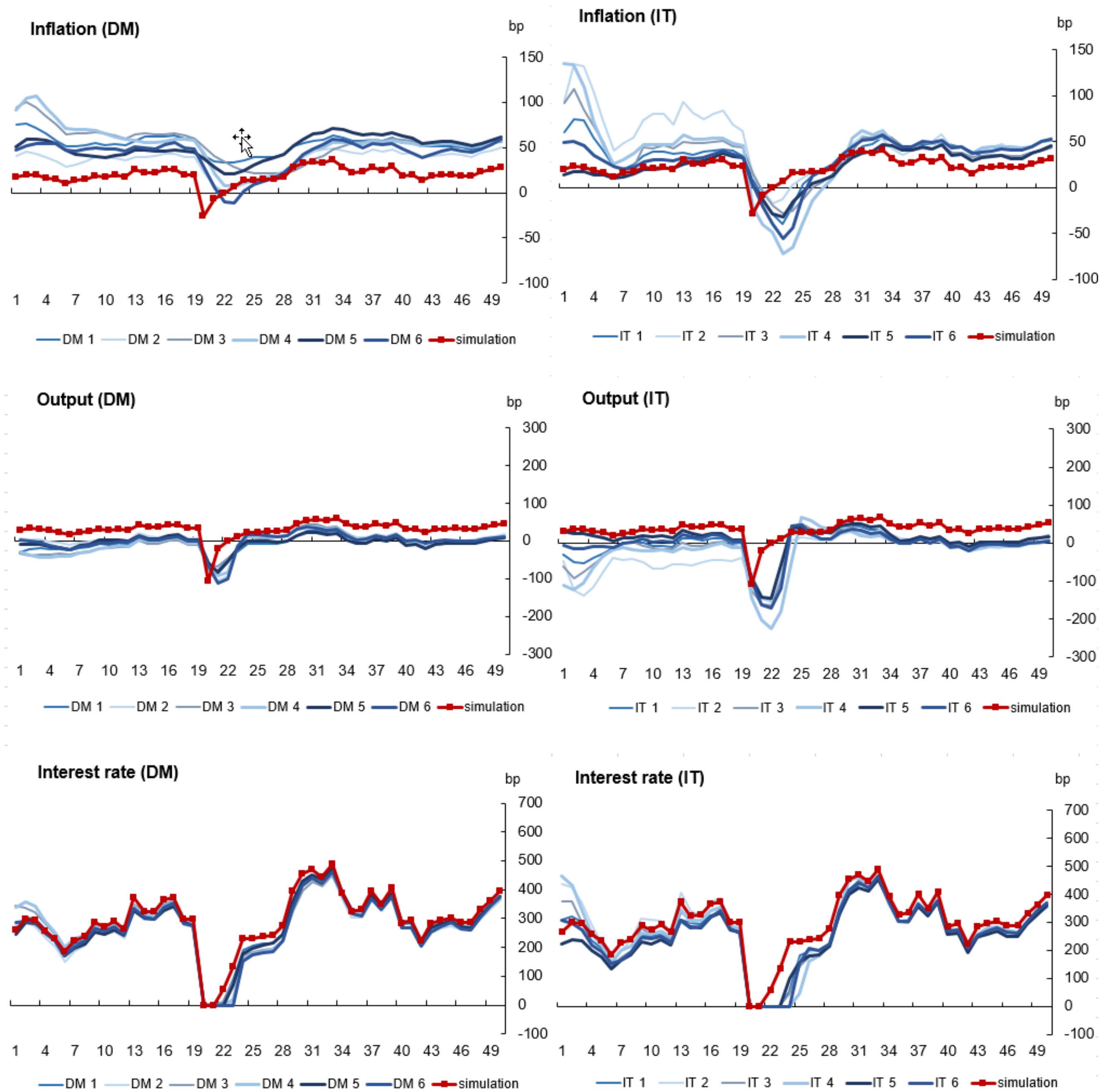


Figure 4: Inflation, output, and interest rate in dual mandate (DM) and inflation targeting (IT) treatments



(a) Dual Mandate (DM) **(b) Inflation targeting (IT)**

Figure 5: Inflation, output, and interest rate in average inflation targeting (AIT) treatment

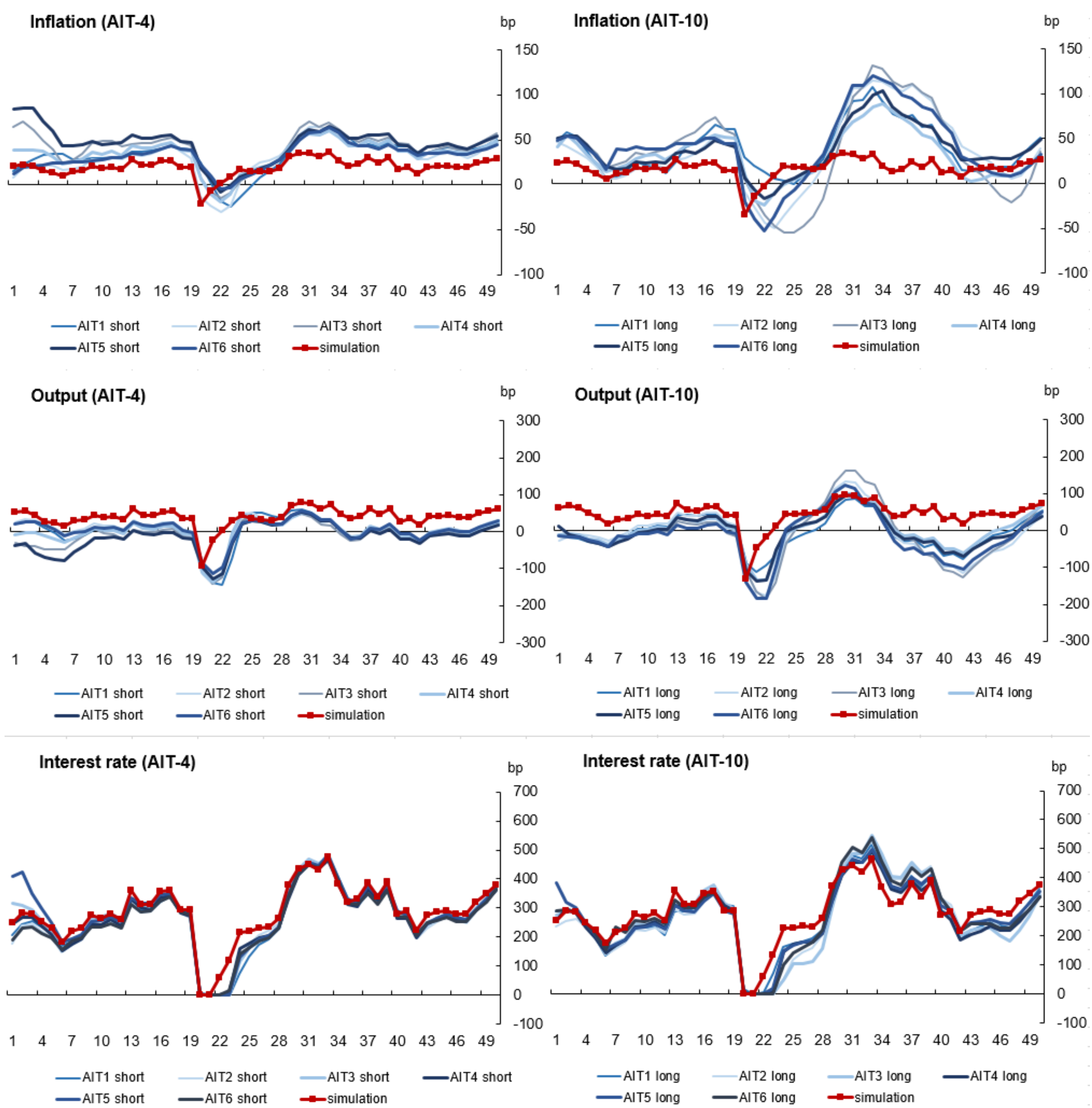
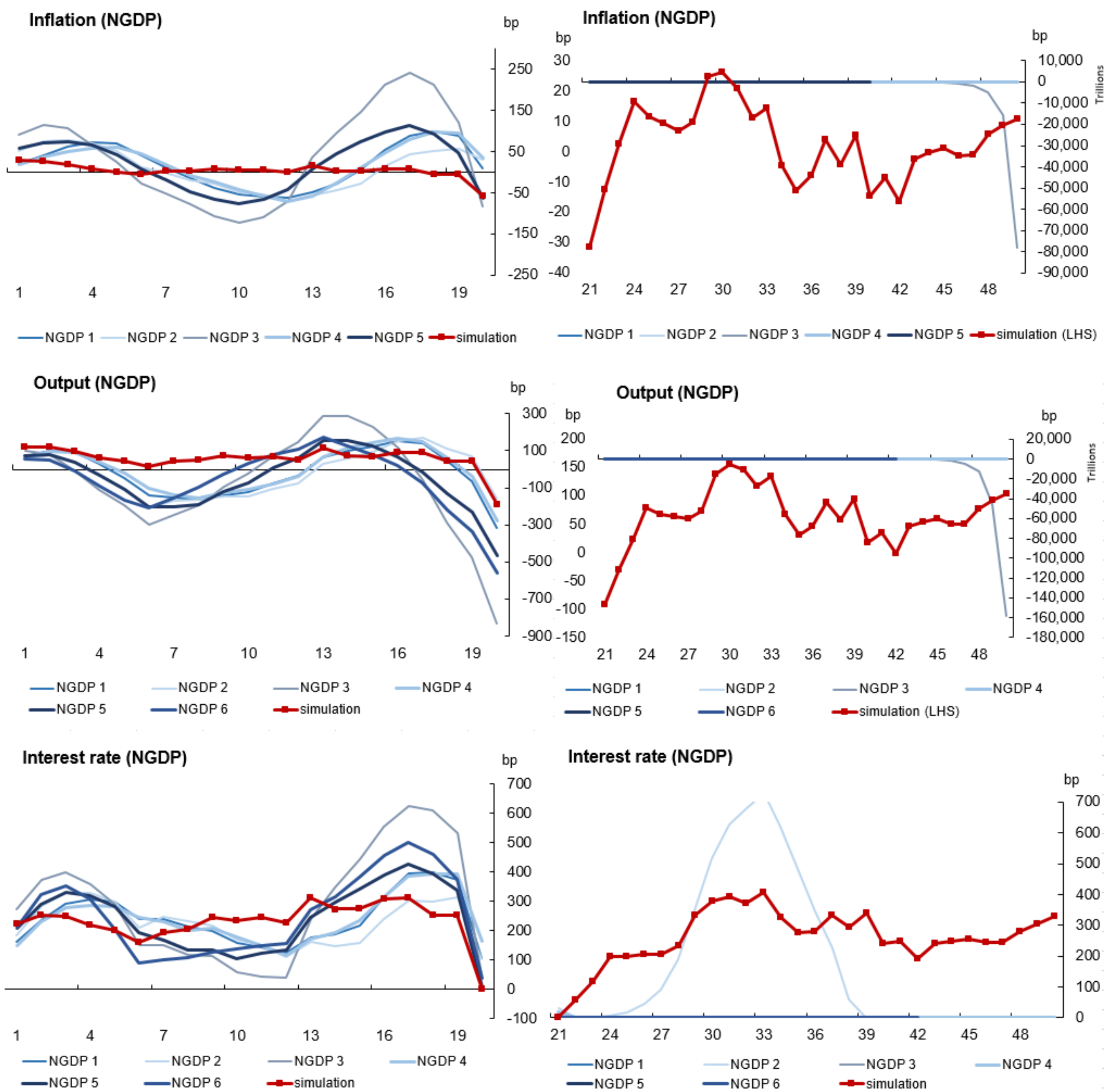


Figure 6: Inflation, output, and interest rate in NGDP level targeting (NGDP) treatment



(a) periods 1-20 (b) periods 21-50

Figure 7: Inflation, output, and interest rate in price-level targeting (PLT) treatment

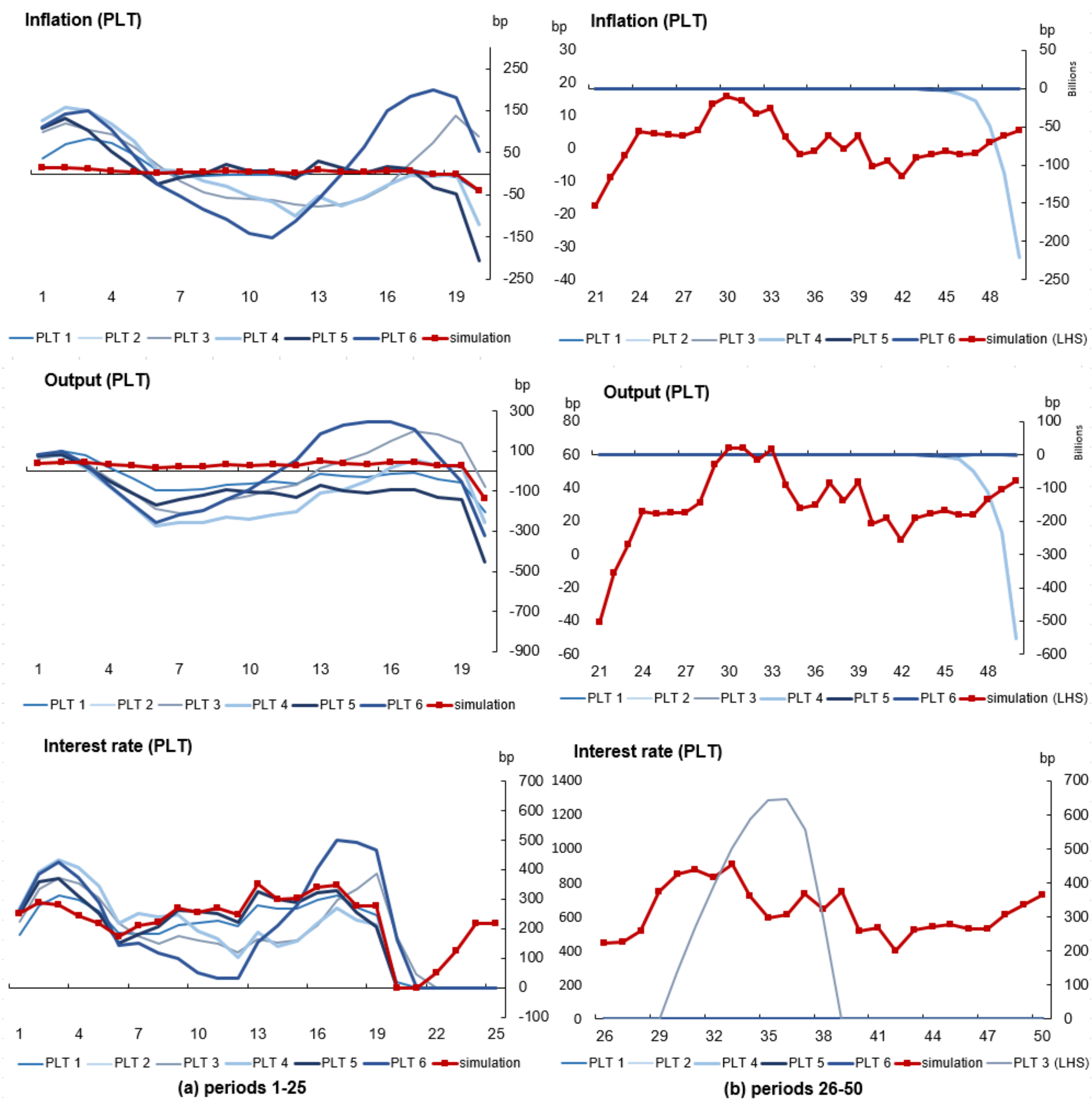


Figure 8: Loss function before shock

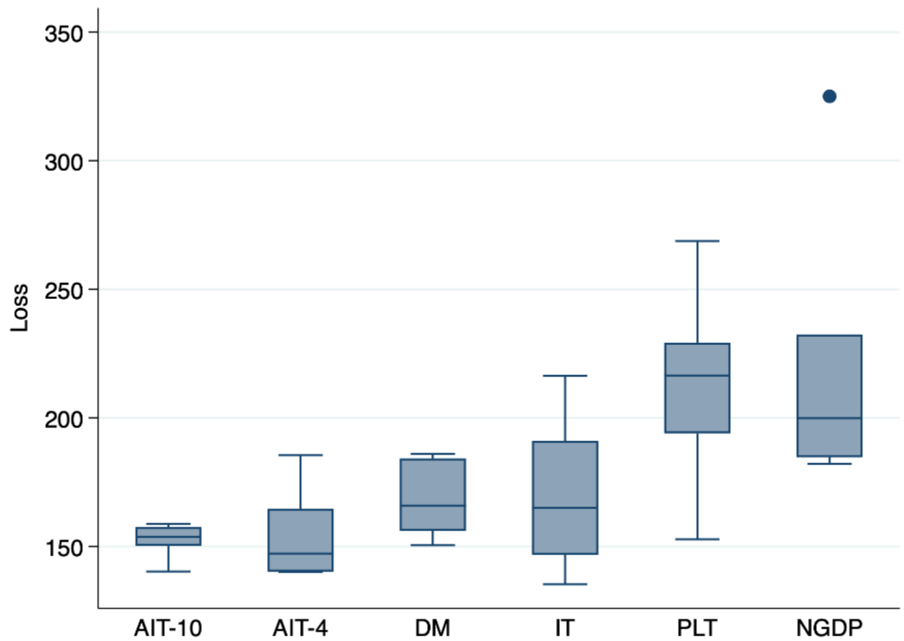


Figure 9: Loss function after shock

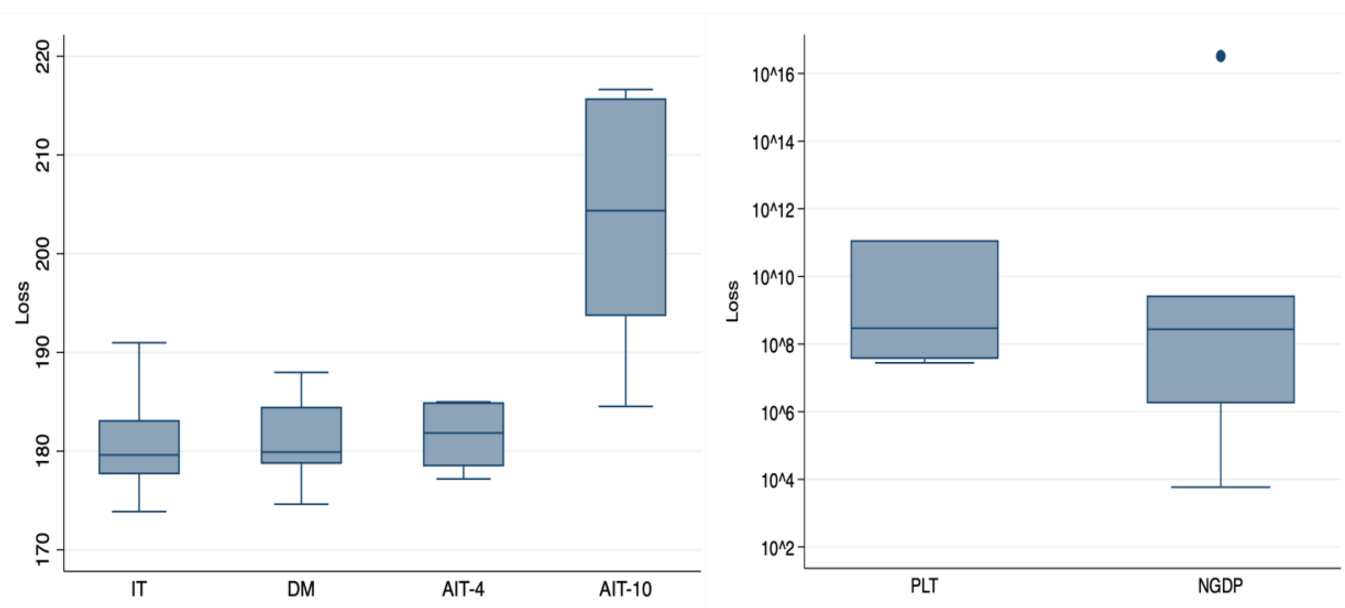
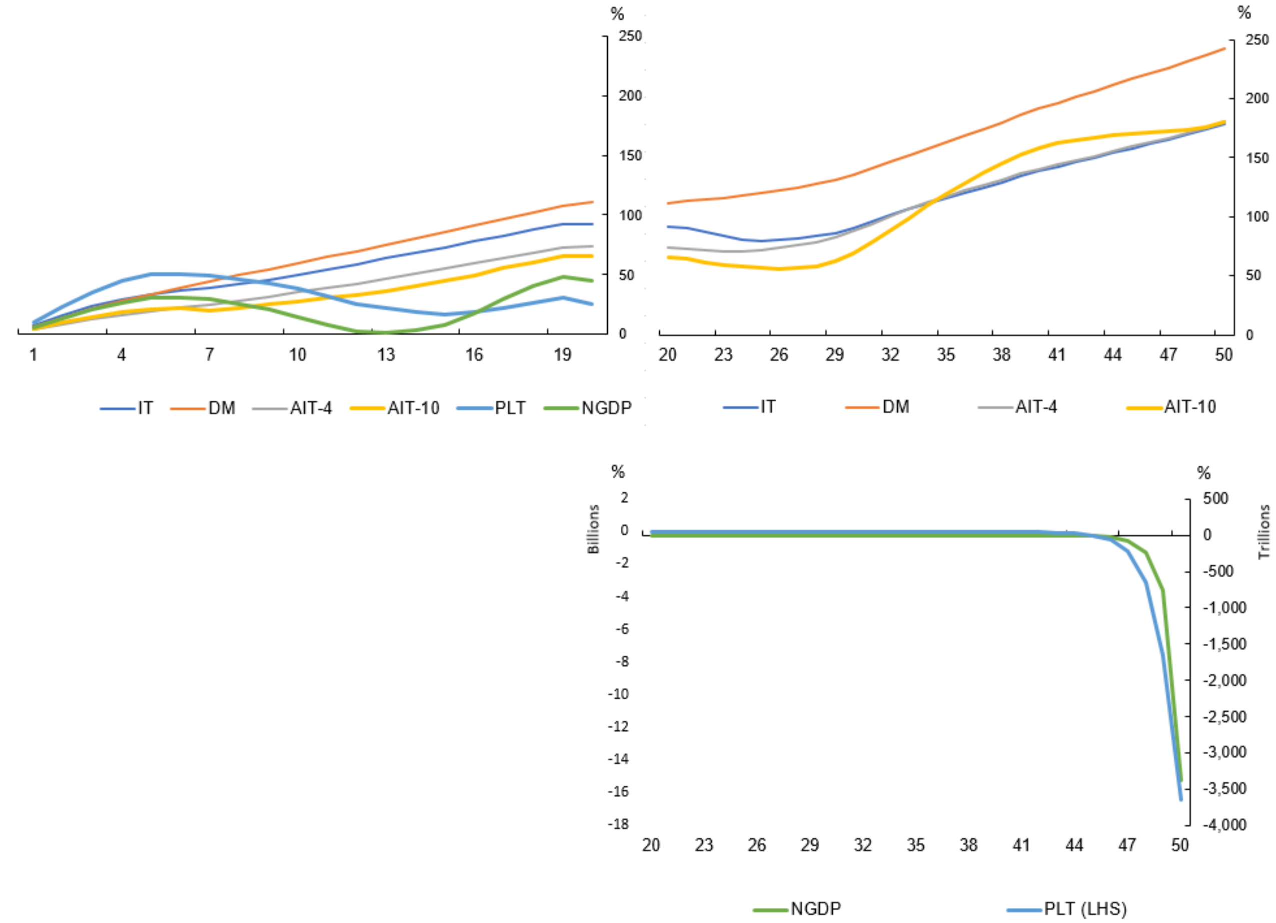
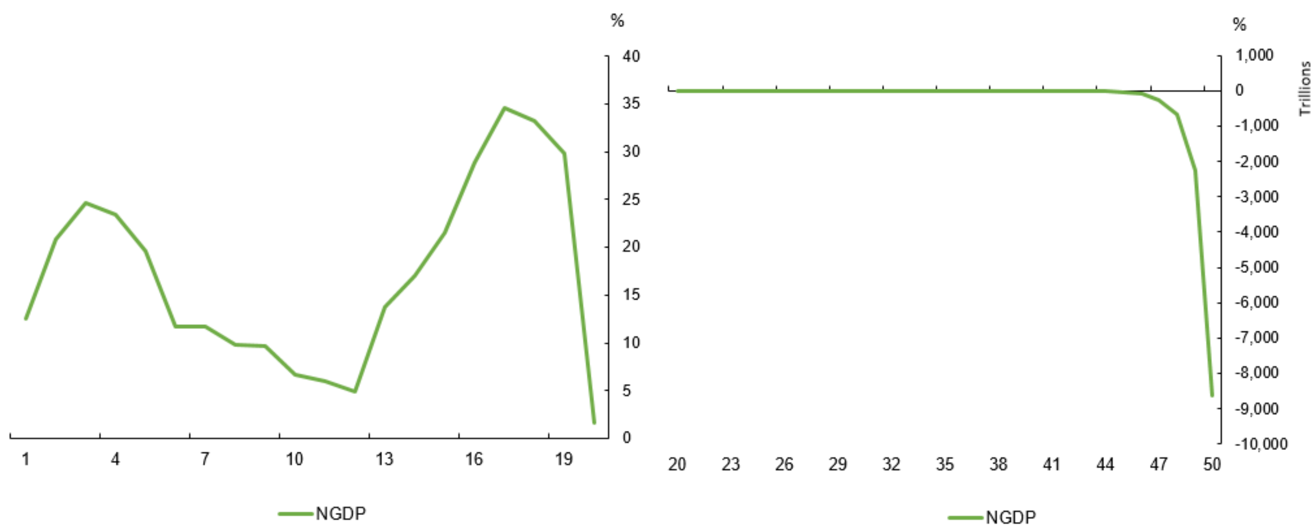


Figure 10: Deviations of price level from price-level target, percent



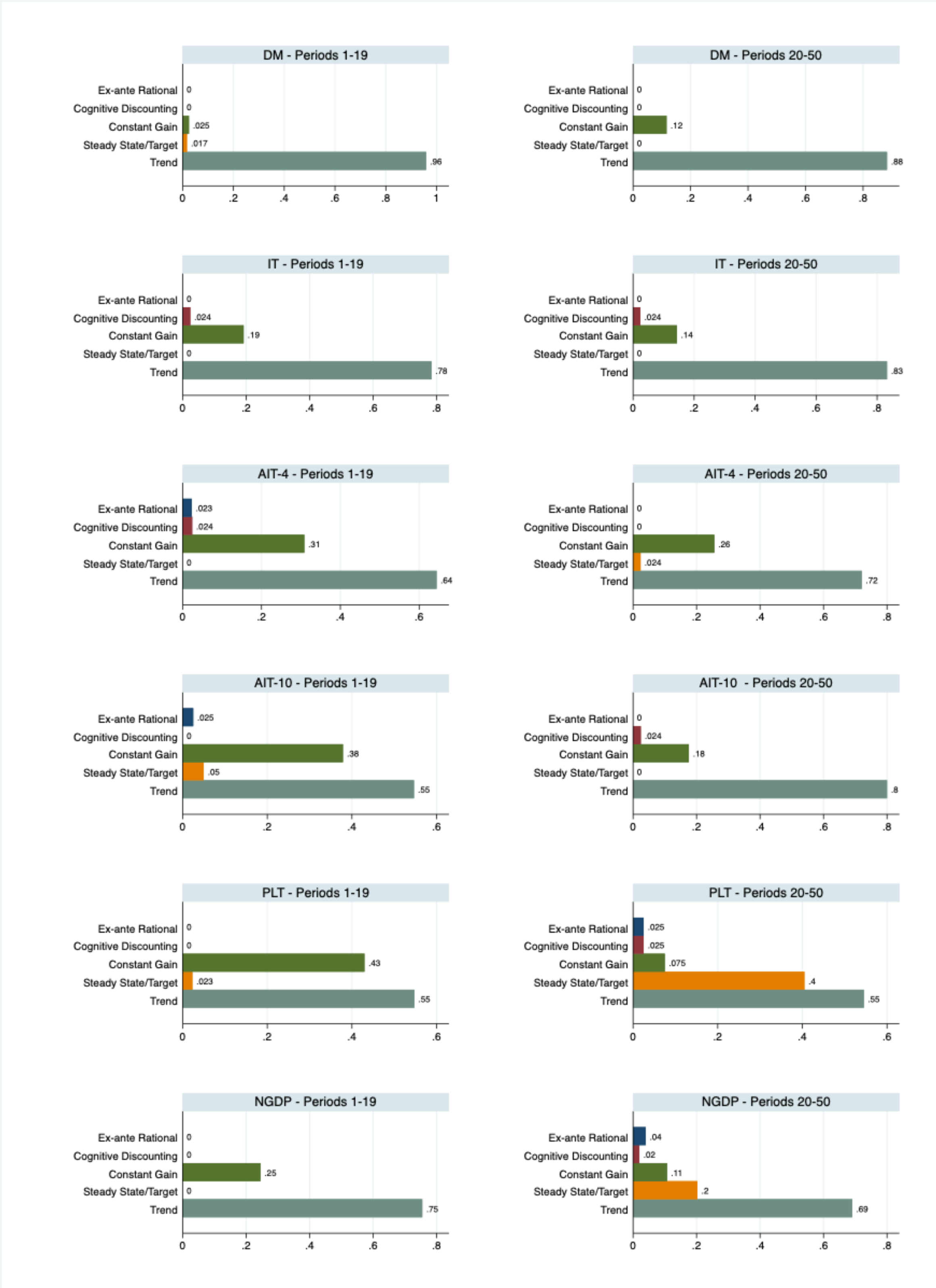
This figure presents deviations of price level from price-level target in PLT and price level consistent with inflation target in other regimes averaged across all sessions of each treatment.

Figure 11: Deviations of nominal GDP from its target, percent



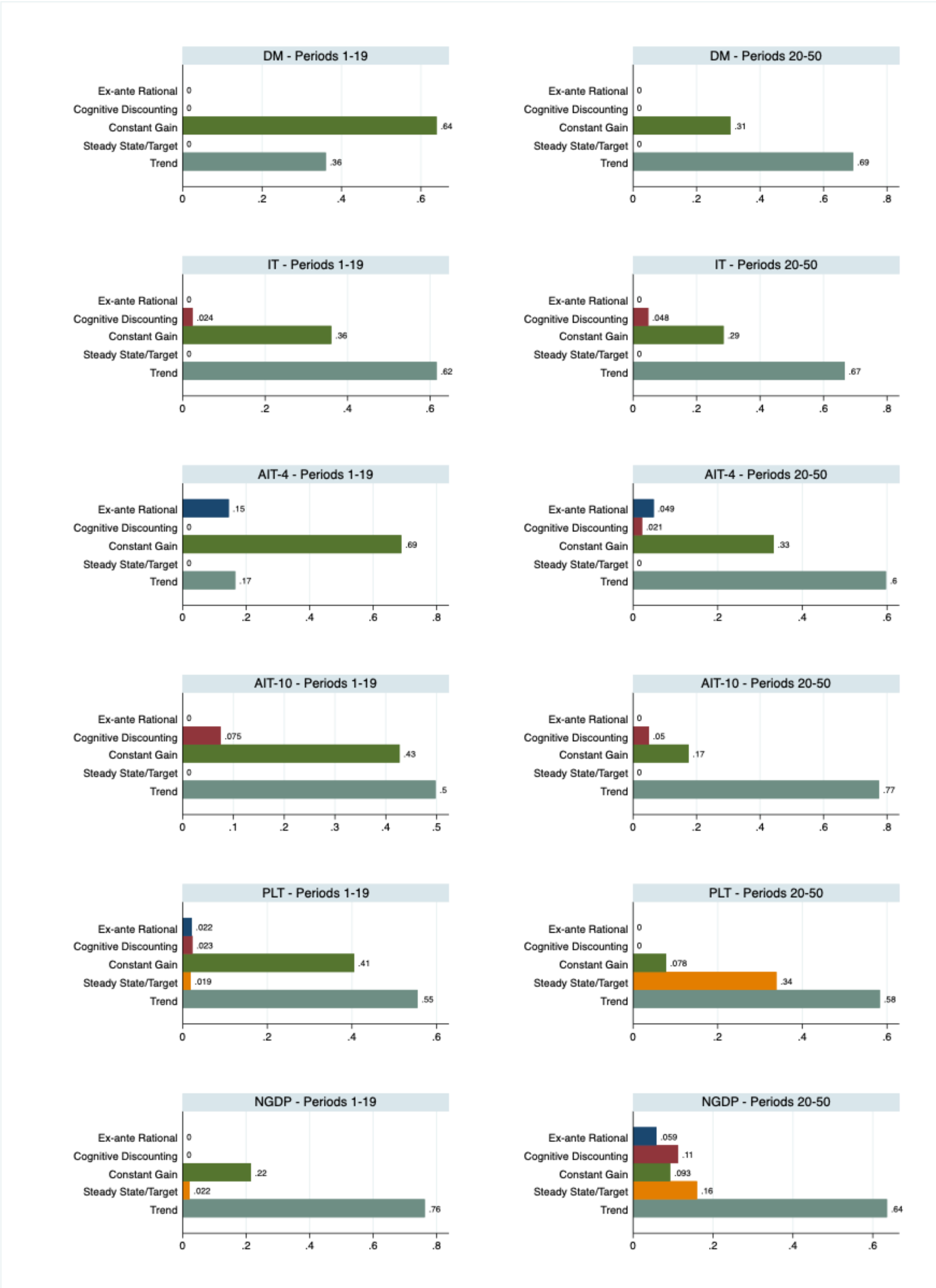
This figure presents deviations of nominal output from its target averages across all sessions in NGDP treatment.

Figure 12: Distribution of forecasting heuristics for inflation forecasts, by treatment and phase



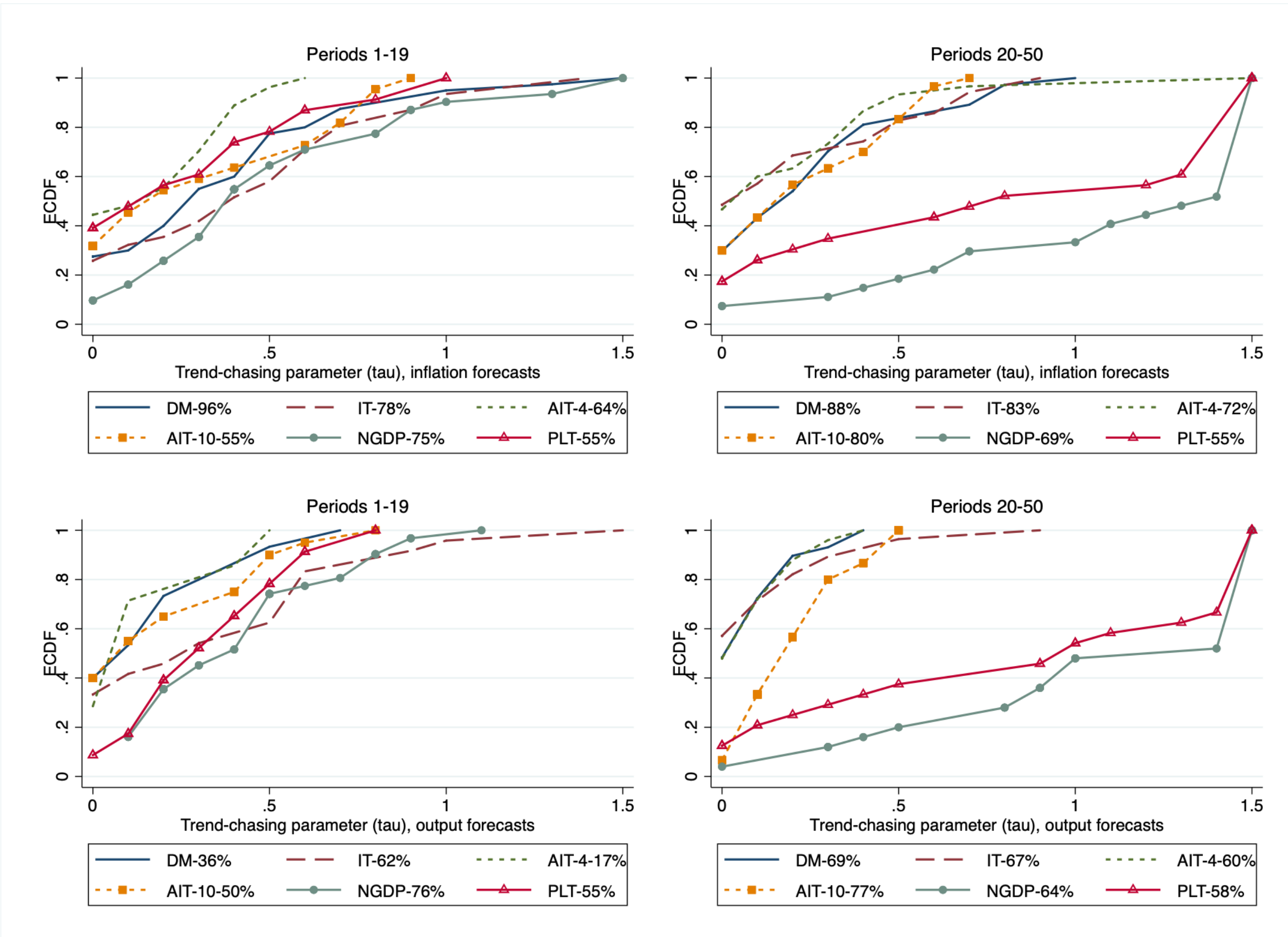
This figure presents the share of participants in each treatment and phase classified into a given heuristic.

Figure 13: Distribution of forecasting heuristics for output forecasts, by treatment and phase



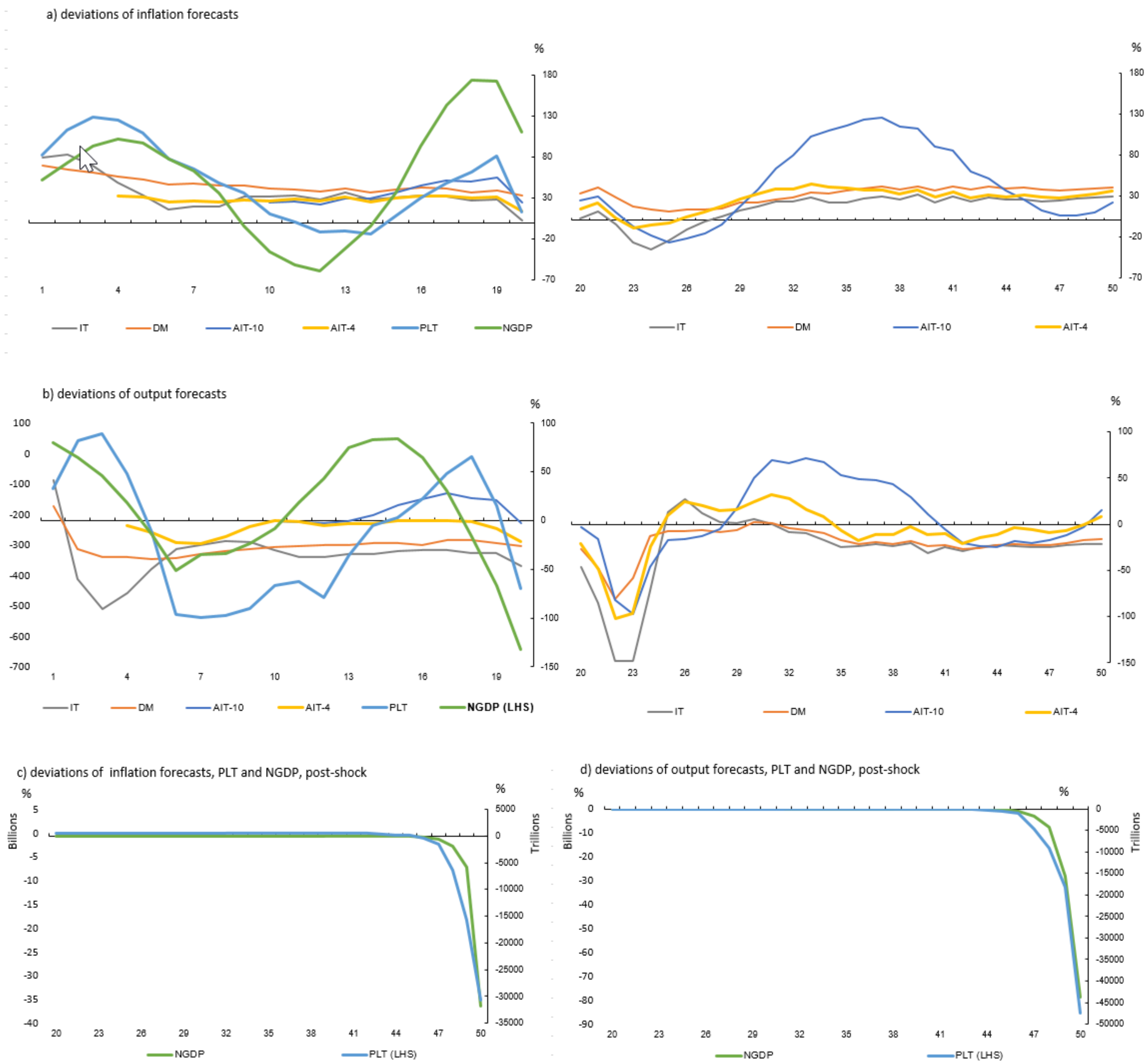
This figure presents the share of participants in each treatment and phase classified into a given heuristic.

Figure 14: Distribution of trend-chasing parameter τ in inflation and output forecasts by phase



This figure presents CDF of parameter τ for participants whose forecasts were classified as trend chasing. Percentages in legend indicate proportion of forecasts classified as trend chasing for corresponding monetary policy regime during the phase depicted in the chart.

Figure 15: Deviations of inflation and output forecasts from REE forecasts

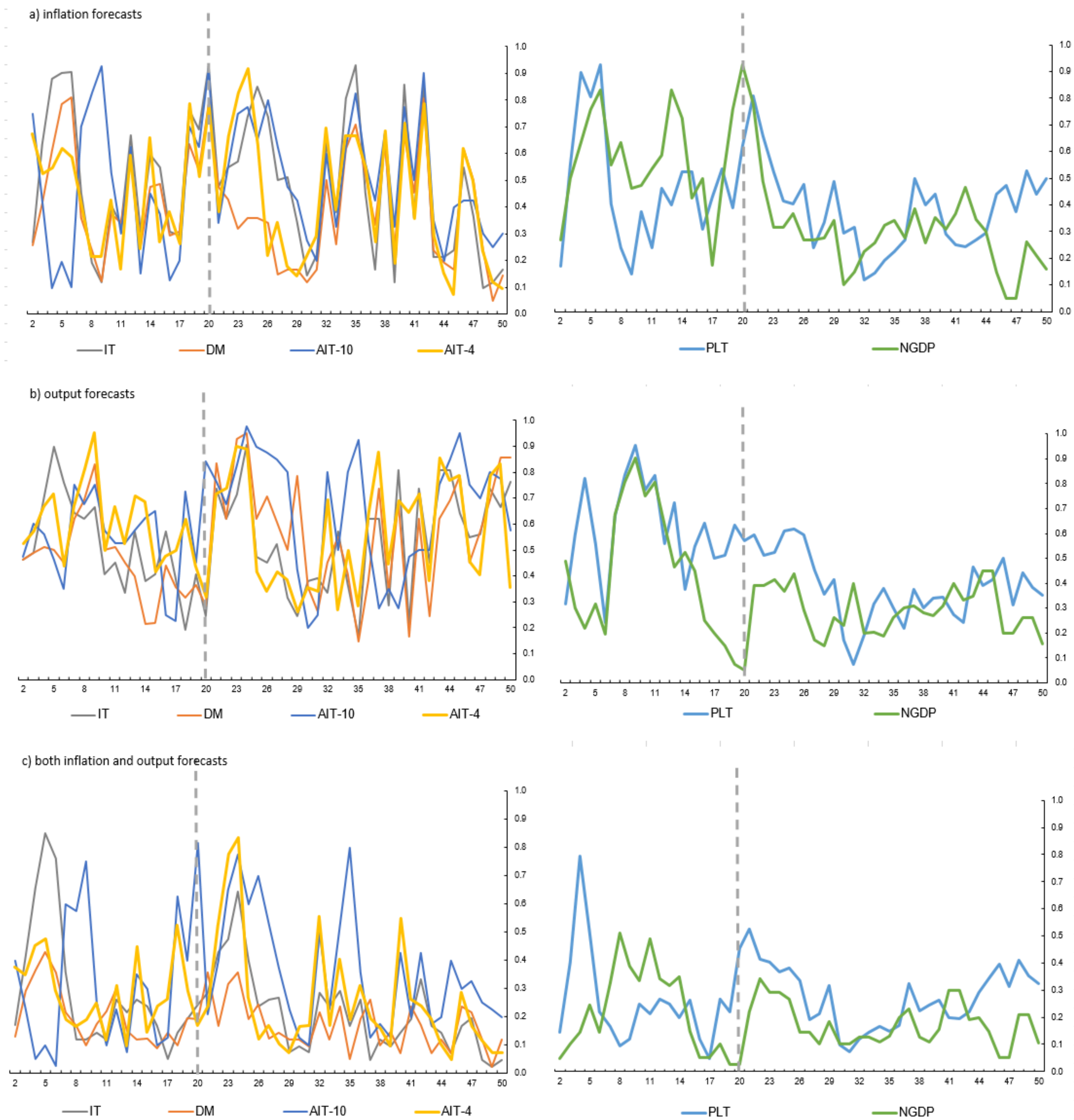


This figure presents the median deviations of inflation and output forecasts from REE, averaged for each period across all sessions for each treatment.

Figure 16: Dispersion of inflation and output forecasts

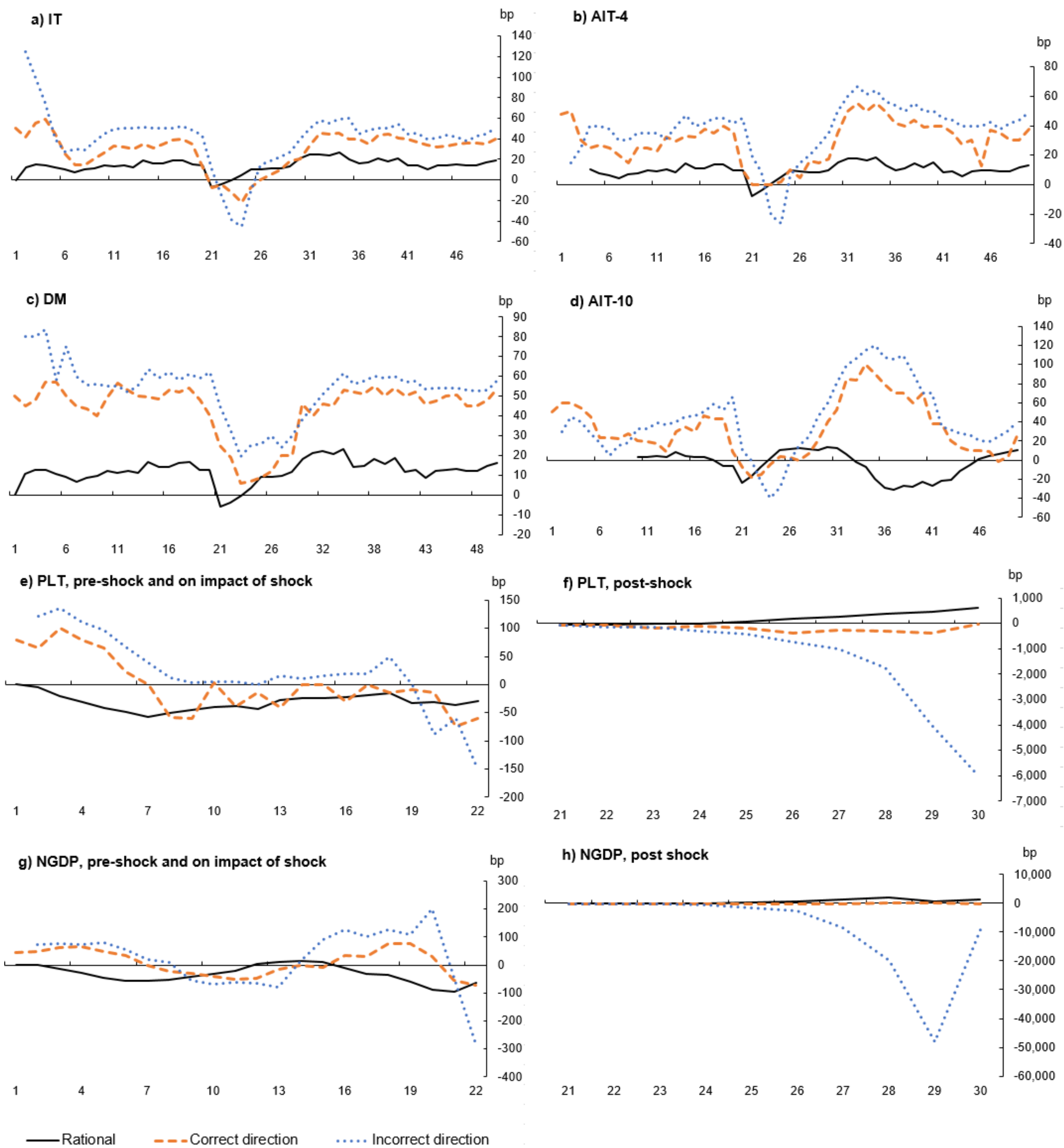
This figure presents the dispersion of inflation and output forecasts as measured by interquartile range, averaged for each period across all sessions of each treatment.

Figure 17: Share of participants exhibiting basic rationality in inflation and output forecasts



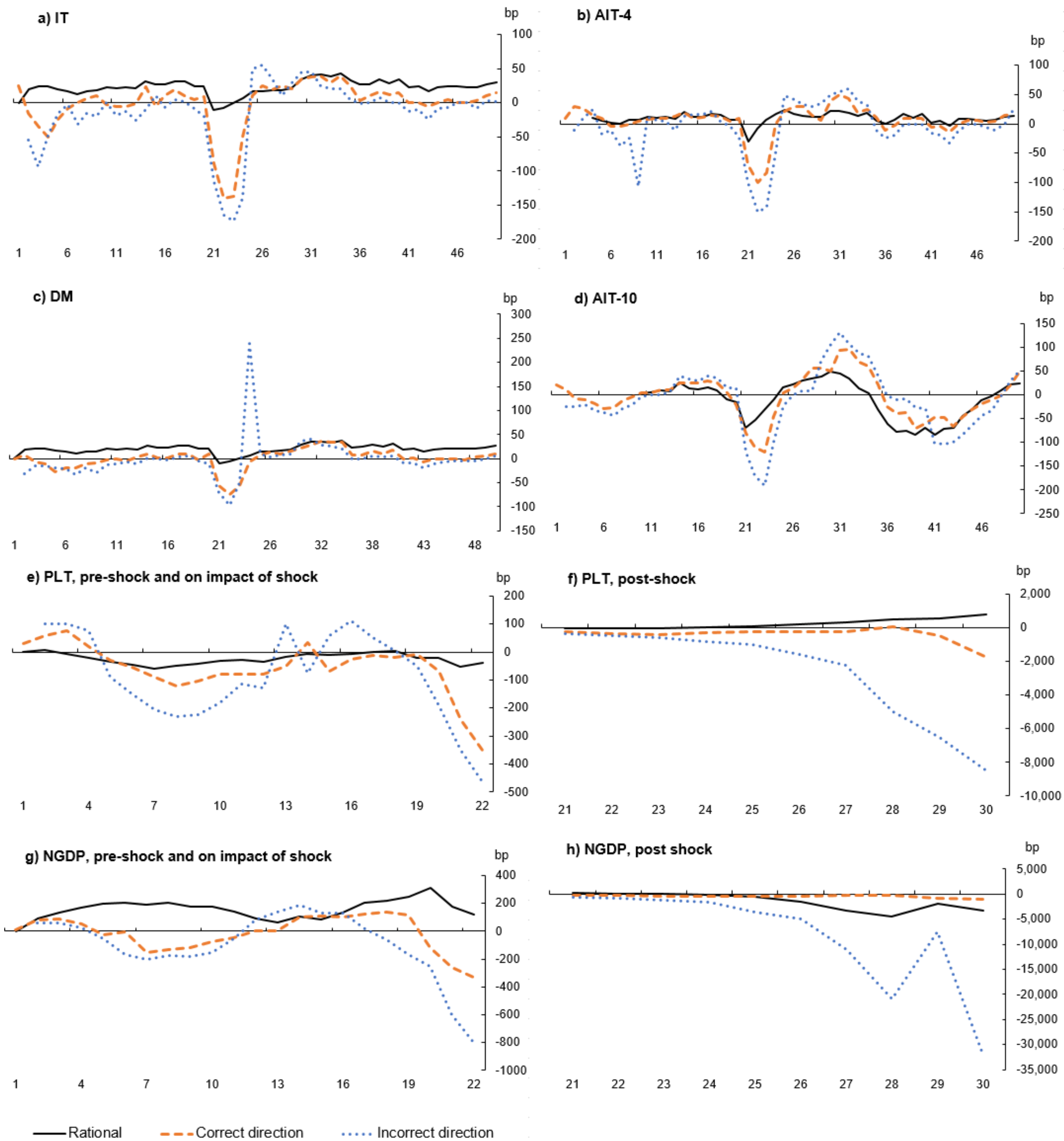
This figure presents the shares of participants whose forecasts satisfy definition of basic rationality as forecasting in the correct direction. Panel (a) presents share for inflation forecasts, panel (b) – share for output forecasts, and panel (c) – share for both inflation and output forecasts satisfying basic rationality.

Figure 18: Inflation forecasts by participants forecasting in the correct direction of rational forecast and incorrect direction



This figure presents the median forecast of participants of each type (correct and incorrect direction), averaged across all sessions of each treatment.

Figure 19: Output forecasts by participants forecasting in the correct direction of rational forecast and incorrect direction



This figure presents the median forecast of participants of each type (correct and incorrect direction), averaged across all sessions of each treatment.

A Appendix

Table A1: Losses associated with inflation, output, and interest rate

	$\sqrt{\text{total loss}/T}$	$\sqrt{\sum^T (\pi - \pi^*)^2/T}$	$\sqrt{\sum^T (x - x^*)^2/T}$	$\sqrt{\sum^T (i - i^*)^2/T}$
periods 1-50				
NGDP	182.5	29.5	87.9	222.3
IT	194.8	34.8	27.5	268.3
DM	195.8	35.3	25.6	270
AIT-4	212	54.9	53.8	279.5
PLT	235.3	63.3	144.3	247.1
AIT-10	247.8	68.5	154.9	255.8
periods 1-19				
NGDP	158.5	17.4	63	204.2
PLT	159.6	11.6	29.2	221.2
AIT-10	173.4	27.8	44	233.9
IT	178.3	31.3	25	245.7
DM	179.2	31.8	23.4	247.2
AIT-4	193.1	48.9	45.9	256
periods 20-50				
NGDP	195.8	34.9	100.2	232.7
IT	204.3	36.8	28.9	281.2
DM	205.3	37.3	26.9	283
AIT-4	222.9	58.2	58	293
PLT	271.4	79.9	181.8	261.6
AIT-10	283.9	84.3	193.7	268.4

Loss and standard deviations of inflation, output, and interest rate were computed from simulations with combination of rational expectations (67%) and naive expectations (33%) and are expressed in basis points. Model becomes unstable when share of naive expectations is above 33% in PLT and above 45% in NGDP targeting. IT, DM, AIT-4, and AIT-10 remain stable even with 100% of naive expectations.

Table A2: Losses associated with inflation, output, and interest rate

	$\sqrt{\text{total loss}/T}$	$\sqrt{\sum^T(\pi-\pi^*)^2/T}$	$\sqrt{\sum^T(x-x^*)^2/T}$	$\sqrt{\sum^T(i-i^*)^2/T}$
periods 1-50				
AIT-10	194.7	37	43.6	263.2
IT	198.4	38.5	23.8	273.2
DM	202.2	42.6	21.1	278
AIT-4	226	69	44.8	297.7
PLT	235.3	63.3	144.3	247.1
NGDP	3010.6	1285.2	2691.2	581.7
periods 1-19				
PLT	159.6	11.6	29.2	221.2
NGDP	161.4	21.6	58	210.9
AIT-10	180.1	35.3	41	242.9
IT	181.4	34.7	21.1	250
DM	184.5	37.7	19.6	253.9
AIT-4	206.8	63	41	272.4
periods 20-50				
AIT-10	203.1	38	45.1	274.9
IT	208.2	40.7	25.3	286.5
DM	212.3	45.3	22	291.7
AIT-4	237	72.5	47	312.2
PLT	271.4	79.9	181.8	261.6
NGDP	3821.4	1632.2	3417.5	720.1

Loss and standard deviations of inflation, output, and interest rate were computed from simulations with a combination of rational expectations and naive expectations and are expressed in basis points. Shares of naive expectations: 33% in PLT and 45% in all other regimes.

Table A3: Losses associated with inflation, output, and interest rate

	$\sqrt{\text{total loss}/T}$	$\sqrt{\sum^T (\pi - \pi^*)^2/T}$	$\sqrt{\sum^T (x - x^*)^2/T}$	$\sqrt{\sum^T (i - i^*)^2/T}$
periods 1-50				
IT	206.7	44.5	28.5	282.6
DM	210.1	49.9	23.5	286.7
AIT-10	217.9	54.1	52	289.3
PLT	235.3	63.3	144.3	247.1
AIT-4	254	95.6	42	327.4
NGDP	3010.6	1285.2	2691.2	581.7
periods 1-19				
PLT	159.6	11.6	29.2	221.2
NGDP	161.4	21.6	58	210.9
DM	183.4	34.2	25	252.3
IT	185.1	35.2	26.4	254.3
AIT-10	200	50	35.8	269.2
AIT-4	211	63.2	46.1	277.1
periods 20-50				
IT	218.9	49.4	29.7	298.6
DM	224.9	57.4	22.6	305.9
AIT-10	228.2	56.5	59.8	301
PLT	271.4	79.9	181.8	261.6
AIT-4	277.1	110.8	39.3	354.8
NGDP	3821.4	1632.2	3417.5	720.1

Loss and standard deviations of inflation, output, and interest rate were computed from simulations with combination of rational expectations and naive expectations and are expressed in basis points. Shares of naive expectations: 33% in PLT, 45% in NGDP, and 100% in IT, DM, AIT-4, and AIT-10.