On the perils of stabilizing prices when agents are learning

Antonio Mele[1], Krisztina Molnár[2] and Sergio Santoro[3]

Abstract

The main advantage of price level stabilization compared with inflation stabilization rests on the central bank's ability to shape expectations. We show that stabilizing prices is no longer optimal when the central bank can shape expectations of agents with incomplete knowledge, who have to learn about the policy implemented. Disinflating in the short run more than agents expect generates short-term gains without triggering an abrupt loss of confidence, because agents update expectations sluggishly. Following this policy, in the long run, the central bank loses the ability to shape agents' beliefs, and the economy converges to a rational expectations equilibrium in which policy does not stabilize prices, economic volatility is high, and agents suffer the corresponding welfare losses. However, these losses are outweighed by short-term gains from the learning phase.

JEL classification: C62, D83, D84, E52

No monetary authority sets price level stabilization[4] as its official goal, despite a vast literature claiming that it is a serious contender as a good way to conduct

[1]University of Surrey; Email: a.mele@surrey.ac.uk

[2]Norwegian School of Economics (NHH) ; Email: krisztina.molnar@nhh.no

[3]European Central Bank, and Bank of Italy; Email: sergio.santoro@bancaditalia.it. A previous draft of this paper has been circulated under the title "The suboptimality of commitment equilibrium when agents are learning". We thank Andrea Caggese, Marco Del Negro, John Duca, Tore Ellingsen, Martin Ellison, Stefano Eusepi, Michal Horvath, Albert Marcet, Ramon Marimon, Andrzej Nowak, Jose Victor Rios Rull, and Aarti Singh for useful comments. All remaining errors are our own. The views expressed herein are those of the authors, and do not necessarily reflect those of the Bank of Italy.

[4]Price level stabilization implies counteracting the effect of shocks on the price level, such that in the long run it reverts to its original value. Hence equilibrium fluctuations in the price level are stationary. In contrast, stabilizing inflation means engineering a stationary inflation but not caring about the absolute level of prices. "Undoing" past deviations in prices would generate unnecessary

monetary policy.[5] This is not because policymakers do not take this recommendation seriously. In fact, Sweden in the 1930s even introduced price level stabilization as the official goal of its monetary policy, after a public debate in which economists supported it.[6] However, this policy was abandoned within the same decade, and today the official goal of Swedish monetary policy is inflation stabilization. More recently, in the aftermath of the 2008 financial crisis, Canada considered introducing long-run price stability as its official monetary policy goal, but decided against it. Policymakers admit that their main concern with this policy recommendation is that the public may have difficulties in understanding it because of its complicated timing and response to shocks.[7] This argument is not about whether the price level is an easier concept to communicate than inflation, but rather, it is about the complexity of price level targeting policies, which agents should understand for its advantages to materialize.

We argue that this concern can indeed rationalize policymakers' reluctance to implement price level stabilization. We show that in a standard macroeconomic model, if there is even a small chance that the private sector could misunderstand the policy regime, then price level stabilization is not optimal.

In our setup, there is a stabilization role for monetary policy, i.e. reducing economic fluctuations by dampening the effect of shocks on aggregate variables. Firms and households know the structure of the economy, but do not perfectly understand

---

fluctuations in inflation, therefore the policymaker "lets bygones be bygones", and the price level is allowed to drift to a permanently different level. (See Woodford [44] Ch 7.)

[5] In particular price level targeting entails history dependence, which turns out to be a robust feature of optimal monetary policy in a wide range of models, see Hatcher and Minford [24].

[6] Swedish economists, such as Gustav Cassel, David Davidson and Eli Heckscher firmly supported price level targeting in public debates, and had a great influence on the government. Knut Wicksell in 1898 was the first in Sweden to present the view that the central bank should aim for price level stabilisation.

[7] This is very transparent in the "Renewal of the Inflation-Control Target" document of the Bank of Canada. The authors write: "[...] these models assume that agents are forward looking, fully conversant with the implications of [price level stabilization] and trust policy-makers to live up to their commitments." (p14.) They argue that it is not clear that these conditions are "sufficiently satisfied in the real world for the Bank to have confidence that price level [stabilization] could improve on the current inflation targeting framework."

how aggregate allocations are impacted by monetary policy. If their understanding were perfect, they could form accurate expectations about how equilibrium allocations depend on shocks. This is the standard rational expectations assumption, and in this case it is a well-established result (see for example Clarida et al. [8] and Vestin [41]) that it is optimal to stabilize prices. The advantage of this policy arises from its history dependence: after a temporary shock that increases the price level, the policymaker engineers a series of aggregate demand contractions in order to bring the price level back to its target; in other words, it can spread out the effect of the shock on the price level through several periods. If agents are aware of this history dependence, the policymaker can lower agents' expectations about future inflation by contracting current output. Lower inflation expectations then decrease current inflation through the Phillips Curve.[8]

We depart slightly from the assumption of rational expectations by postulating that even if agents knew that aggregate variables depend on shocks, they do not know the exact mapping induced by monetary policy.[9] We assume that agents learn the mapping between shocks and aggregate variables by extrapolating from historical patterns in observed data. More specifically, they rely on econometric methods to estimate a model of the economy and use it for forecasting future aggregate variables. In each period, as new observations are available, they update their model in order to have more precise beliefs. Therefore, they have a chance to learn the exact mapping (i.e., one that is consistent with rational expectations beliefs), provided they can collect enough data.

Our paper develops further the literature featuring a rational policymaker that behaves optimally when the private sector does not have rational expectations. Like

[8]Our model uses a sticky price framework. Inflation depends on inflation expectations because firms know they might not be able to reset their price in the future, and therefore they must be forward looking when setting their price.

[9]We find this assumption an appealing way to introduce agents' misunderstanding in an otherwise standard model. Agents' knowledge of their own optimization problem does not imply they can derive aggregate allocations that arise in equilibrium (Adam and Marcet [1]). Moreover, an individual might be uncertain about other agents' knowledge about the exact mapping, which in turn would impact the evolution of aggregate variables.

Gaspar et al. [21] and Molnar and Santoro [31] we consider a central bank that takes into account how its policy actions affect the data used in agents' estimations, and how those data affect their future beliefs.[10] Our main contribution with respect to their treatment is that the model of the economy estimated by the private sector is general enough to nest two different mappings, one consistent with price level stabilization and the other with inflation stabilization, while in their analysis it nested only the latter.

This generalization has important implications for the policy design, which now features an equilibrium selection problem. In our setup the monetary authority can "teach" agents either of the two mappings: by choosing a particular policy response to shocks, the policymaker affects agents' beliefs about the mapping, which in turn feed back into the evolution of aggregate variables, and thus into the mapping between shocks and aggregate variables. Hence, differently from the previous papers, agents can in principle learn price level stabilization, which is considered in the rational expectations literature the best policy to implement.

As such, we refine the existing concept of learnability. Several authors have suggested that learning can be used for equilibrium selection, and examined how policy can guarantee a learnable equilibrium (see Evans and Honkapohja [16] and Eusepi and Preston [13] for extensive surveys). Our paper extends their analysis taking into account strategic interaction between a large, rational player and learning agents. We think that this extension of the policy problem with equilibrium selection is appealing when there is a theoretical possibility of teaching different rational expectations equilibria.

Our main result is that price level stabilization is no longer optimal, despite being feasible. This is a strong result, given that the policymaker could induce agents to learn stable prices, and anchor their expectations, but instead gives up the benefits of stabilizing the price level in favor of short-term gains.

Under learning the CB can attain short-term gains because agents revise their

[10] Eusepi et al. [12] derive the optimal long-run inflation rate in a New Keynesian model extended to account for a low-frequency drift in beliefs.

beliefs sluggishly. We show that under learning it is optimal to contract current output very aggressively, instead of spreading out the output contractions over several periods. The policymaker can do this because agents need to gather sufficient data to discover that the policy has become less history dependent. In the meantime the policymaker can still anchor inflation expectations, and lower current inflation by contracting output.

These CB incentives arise due to a fundamental difference between learners and rational agents. Deviation from the price stabilizing policy would be immediately realized by rational agents, who in turn would change their beliefs abruptly and infer that the central bank is following an alternative policy. This off-equilibrium threat of rational agents can keep the CB from deviating from the price stabilizing policy (see Kurozumi [26]). In contrast, adaptive learners do not have separate off-equilibrium strategies. They only learn from realized outcomes, and their strategies are the same with a deviating and not-deviating CB. This lack of off-equilibrium strategies provides strong incentives for the rational policymaker to deviate from the price stabilization policy.

In the long run, monetary policy completely loses its ability to engineer a history-dependent policy that could anchor agents' inflation expectations, because agents eventually learn that the policymaker is not implementing a price level stabilization policy. This policy can be described as *stabilizing inflation instead of the price level*: the CB responds to shocks as long as they affect inflation. The long-run policy recommendation is therefore in line with what many CBs set as their official goal.

What makes our result compelling is that the transition matters for the long run equilibrium; policy incentives during the transition inform the long-run behavior of optimal policy. The long-run benefit of anchoring prices has already been established in the literature, and under learning the mechanism is the same as under rational expectations, namely expectations are better anchored. The CB could attain price level stabilization in the long run simply by implementing it long enough. Yet, it is optimal to drive the economy away from stabilizing prices, because during the transition short-run policy incentives generate high welfare gains.

The policymaker has no incentive to build credibility (in the sense that it can anchor inflation expectations by contracting output). Along the transition, as long as the CB has some credibility, it also has an incentive to exploit it. In the long run, when agents learn to ignore output contractions in forming their inflation expectations, temporarily revamping even little credibility becomes too costly for the CB, especially because it would lose it immediately.

In our framework, the standard assumptions for proving convergence commonly used in the learning literature are not satisfied. This complication arises because of the interaction between atomistic learning agents and a rational strategic player (the CB), which the previous literature did not consider. We therefore derive a novel convergence theorem that can accommodate the interaction between updating rules for agents' beliefs and the choices of the rational CB. This methodological contribution might be of separate interest to some readers, as our theorem and our line of proof could be applied to similar problems with a linear-quadratic setup.

Our paper adds a new insight to the debate about price level targeting (PLT) without questioning its long-run benefits. We show the presence of new short-run policy incentives that can counterbalance long-run benefits of PLT when there is even a small chance that agents could misunderstand policy choices. In our setup it is not optimal to preserve those advantages of PLT that rest on the policymaker being able to use history-dependent policy to influence future beliefs.[11] This history dependence was previously proven to be robust along several dimensions (for example output uncertainty in Gorodnichenko and Shapiro [23], and model uncertainty in Aoki and Nikolov [2]),[12]

We present the model in Section 1 and solve it in Section 2. We derive optimal policy in Section 3 and discuss how to approximate it with a simple rule in Section 4. In Section 5 we relax our main assumptions; finally, Section 6 presents concluding remarks.

---

[11] For a neat summary of the advantages of targeting prices and its practical aspects see Reis [35].

[12] PLT can also alleviate the risks of hitting the zero lower bound (Eggertsson and Woodford [11], Wolman [42]). In some extensions of the baseline model a base-level drift of the price level is optimal, for example when firms are indexing to past inflation, see Røisland [37].

1. The Model

We develop our idea by weakening the assumption on private sector expectations in the well-known monetary policy analysis of Clarida et al. [8]. This example is chosen because the policy implications under rational expectations are well-known to many readers.

The CB can bring about any evolution of inflation π_t, output gap x_t and nominal interest rate r_t, consistent with the aggregate demand and supply equations

$$x_t = E_t^* x_{t+1} - \sigma^{-1}(r_t - E_t^* \pi_{t+1}), \tag{1}$$

$$\pi_t = \beta E_t^* \pi_{t+1} + \kappa x_t + u_t, \tag{2}$$

where $\sigma > 0$, $0 < \beta < 1$, and $\kappa > 0$.[13] The cost-push shock is $u_t \sim N(0, \sigma_u^2)$.[14] E_t^* denotes conditional expectations of the private sector, which are not necessarily rational. The analysis is simplified by assuming that agents have common expectations, and have common knowledge about this; given this the linear aggregate relations can be derived with the usual log-linear approximation to equilibrium relations.[15]

The CB seeks to minimize a quadratic loss function [16]

[13] σ is the household's risk aversion parameter, β denotes the subjective discount rate, and κ is a function of structural parameters. For details of the derivation of the structural equations of the New Keynesian model see, among others, Yun [46] and Woodford [44].

[14] This assumption is supported by Milani [28] who estimates an i.i.d cost-push shock in the presence of learning. It also makes the problem more tractable, and allows us to focus on the policy implications of nonrational beliefs.

[15] As pointed out by Preston [32], for arbitrary nonrational beliefs satisfying standard probability laws, the solutions to intertemporal optimization problems require agents to make infinite horizon forecasts. Here, following Honkapohja et al. [25] we assume that agents understand that other agents have the same tastes and beliefs; therefore, the law of iterated expectations holds and we can simplify intertemporal decisions to one-step-ahead forecasts about their payoff-relevant variables.

[16] The period loss function is derived as a quadratic approximation to household utility. The derivation is not affected by assuming nonrational expectations. For the derivation see Rotemberg and Woodford [38] and Woodford [44]. The parameter α is a function of structural parameters. The optimal output gap is zero, as distortions from firms' monopolistic competition are assumed to be corrected with an appropriate labor cost subsidy.

$$E_0(1-\beta)\sum_{t=0}^{\infty}\beta^t\left(\pi_t^2+\alpha x_t^2\right), \tag{3}$$

where $\alpha > 0$. Here the policymaker is considering the effects of alternative policies, and E_0 denotes conditional expectation based on CB beliefs. We focus on a rational CB that knows the structure of the economy, including how agents form their expectations, which allows us to gauge how a learning private sector changes incentives for monetary policymaking.[17]

The novelty of this setup is that the policymaker can drive agents to certain equilibria in the long run (Section 2) and also affects how they should learn during the transition (Section 3). In fact, early literature on adaptive learning motivated it as a way to select amongst multiple rational expectations equilibria. In our setup, learnability of an equilibrium is not sufficient for it to arise in the long-run; the strategic behavior of the rational policymaker can affect the optimal long-run equilibrium. It is undoubtedly a strong assumption that the CB knows how agents form their expectations; we relax this in Section 5.

1.1. Price level targeting vs inflation targeting under RE

When the agents are rational and the CB can credibly commit to future policy, optimal allocations have the following law of motion[18]:

$$x_t = b^x x_{t-1} + c^x u_t, \tag{4}$$

$$\pi_t = b^\pi x_{t-1} + c^\pi u_t, \tag{5}$$

where $b^x = \frac{\kappa^2+\alpha(1+\beta)-\sqrt{(\kappa^2+\alpha(1+\beta))^2-4\alpha^2\beta}}{2\alpha\beta}$, $c^x = -\frac{\kappa b^x}{\alpha}$ and $b^\pi = \frac{\alpha}{\kappa}\left(1-b^x\right), c^\pi = -\frac{\alpha}{\kappa}c^x$. This policy is equivalent to PLT: the CB responds to changes in the price level, and tries to keep prices close to a predetermined value. In equilibrium the price

[17] Because the CB and the agents form expectations in different ways, the CB is not a benevolent planner, and it does not maximize the expected utility perceived by agents.

[18] See Clarida et al. [8] and Vestin [41].

level follows a stationary process.[19] The advantage of price level stabilisation arises from its history dependence: in a forward-looking environment history dependence entails welfare gains, because the policymaker can lower agents' expectations about future inflation by contracting current output and spreading the cost of adjustment to shocks over several periods. This history dependence is a robust feature of the optimal policy, even in setups more complicated than ours (see Hatcher and Minford [24]).

When the CB cannot commit to future policy the optimal allocations are

$$x_t = -\frac{\kappa}{\alpha + \kappa^2} u_t \tag{6}$$

$$\pi_t = \frac{\alpha}{\alpha + \kappa^2} u_t \tag{7}$$

We call this inflation targeting (IT in short), because Clarida et al. [8] show that the CB responds to changes in inflation, by trying to stabilize the inflation rate.

These policies differ in a crucial respect. The PLT policy is an inertial policy in the sense of Woodford [43]: the current allocations depend on past levels of output gap. On the contrary, the IT policy only depends on current shocks.

1.2. Learning specification

In the remainder of the paper, we assume that agents are adaptive learners: they know their own optimization problem, observe aggregate variables and prices that are exogenous to their decision problem, and know that other agents are identical to them.[20] However, based on the internal rationality concept of Adam and Marcet [1] we assume that agents' knowledge of their own optimization problem does not imply they can derive aggregate allocations that arise in equilibrium. Our agents have an imperfect understanding of the prevailing policy regime, therefore even though they are able to calculate the rational expectations equilibrium, they are uncertain about

---

[19] The equilibrium price level consistent with (4)-(5) is $p_t = \delta p_{t-1} + \delta u_t$, where $\delta \equiv (1 - \sqrt{1 - 4\beta\gamma})/(2\gamma\beta) \in (0, 1)$, and $\gamma \equiv \alpha/(\alpha(1 + \beta) + \kappa^2)$.

[20] See Preston [32] on infinite horizon learning that results when agents do not know others are identical.

the values of its parameters', and estimate these adaptively by observing past and current allocations.

More precisely, we assume that agents do not know the exact process followed by the endogenous variables, but recursively estimate a Perceived Law of Motion (PLM) consistent with the law of motion that they would observe if the CB followed the PLT policy under RE:[21]

$$\pi_t = b^\pi x_{t-1} + c^\pi u_t \tag{8}$$
$$x_t = b^x x_{t-1} + c^x u_t, \tag{9}$$

Under learning, agents estimate the coefficients in equations (8)-(9), and use their estimates of b^π_{t-1} and b^x_{t-1} and the i.i.d. nature of u_t to make forecasts[22]:

$$E^*_t \pi_{t+1} = b^\pi_{t-1} x_t, \quad E^*_t x_{t+1} = b^x_{t-1} x_t \tag{10}$$

A novel feature of (8)-(9) is that private expectations are consistent with both PLT and IT; hence, agents can learn both those policies, depending on the policy followed by the CB.

At time t, the CB can impact private beliefs by engineering current output contractions or expansions. This makes a nice parallel to the case of CB credibility under rational private beliefs: current actions of the CB impact private beliefs immediately, as long as agents believe the CB can do so, i.e. as long as b^x, b^π are bounded away from zero. Whereas under rational private beliefs a CB could also make promises about the future, under learning this is not possible. Rational agents would be able to think forward, thus promises of future output contractions impact current beliefs,

[21] Agents could make use of more variables to make their forecasts or use an underparameterized model. In the former case, depending on the CB policy, they could learn the RE equilibrium, while in the latter case it is clear that they cannot. Although these scenarios are of interest, they are beyond the scope of this paper.

[22] Agents forecast self-referential variables, i.e. ones that depend on the agents' actions. In this kind of models a rational Bayesian learner's expectation has not yet been solved: she would understand how her actions impact on the variable in question, and would not treat the posterior as random, but instead would have to calculate the posterior as a complicated fixed point problem. This makes adaptive learning especially useful, because agents simply infer from past allocations.

as long as they are credible. Under learning, on the other hand, the impact of an output contraction depends solely on the learning coefficients b^x, b^π, which in turn depend on the history of past CB actions.[23]

We assume that agents' estimates are obtained with stochastic gradient learning (SG) (Barucci and Landi [3] and Evans and Honkapohja [15]), which is a plausible learning device from a bounded rationality standpoint, because it keeps the state space small by abstracting from the evolution of the estimated second moments of the regressors.[24] The recursive updating formula for the estimated coefficients is

$$\begin{aligned} b_t^\pi &= b_{t-1}^\pi + \gamma_t x_{t-1} \left(\pi_t - x_{t-1} b_{t-1}^\pi\right) && (11) \\ b_t^x &= b_{t-1}^x + \gamma_t x_{t-1} \left(x_t - x_{t-1} b_{t-1}^x\right), && (12) \end{aligned}$$

where γ_t is the so-called gain parameter, determining the rate at which older observations are discounted. When deriving our analytical results, we use $\gamma_t = \frac{1}{t}$ (decreasing gain learning). As t increases $\frac{1}{t} \to 0$, agents perceive all changes as temporary. This allows us to establish convergence to a nonstochastic point as t increases.[25]

The timing is as follows. At each period t agents inherit belief parameters b_{t-1}^π, b_{t-1}^x, determined by period $t-1$ data. They use their forecast function (10) to form expectations about future variables. Agents use (11) to update the coefficient estimates b_t^π, b_t^x, based on their inherited coefficients b_{t-1}^π, b_{t-1}^x and new data π_t, x_t. In the spirit of anticipated utility (Sargent [39]), agents do not take into account that their beliefs will be updated in subsequent periods, and forecast as if their forecasting coefficients were fixed.

[23] An alternative timing assumption is when agents cannot observe contemporaneous x_t, which would limit the CB's ability to impact private beliefs.

[24] This assumption also delivers analytical tractability with the new convergence theorem, which we present in the next section.

[25] As shown in Evans and Honkapohja [16], with a small constant γ, beliefs would be ergodically distributed around the convergence point.

2. Optimal monetary policy

Following Molnar and Santoro [31], we posit that the CB is fully rational, it knows the structural equations that characterize the economy, and how private agents form and revise their beliefs; hence, it solves the following problem:

$$\sup_{\{x_t, b_t^\pi, b_t^x\}_{t=0}^\infty} E_0(1-\beta)\sum_{t=0}^{\infty}\beta^t\left\{-\frac{1}{2}\left[\left((\beta b_{t-1}^\pi+\kappa)x_t+u_t\right)^2+\alpha x_t^2\right]\right\} \tag{13}$$

s.t.

$$b_t^\pi = b_{t-1}^\pi + \gamma_t x_{t-1}\left((\beta b_{t-1}^\pi+\kappa)x_t+u_t-x_{t-1}b_{t-1}^\pi\right) \tag{14}$$

$$b_t^x = b_{t-1}^x + \gamma_t x_{t-1}\left(x_t - x_{t-1}b_{t-1}^x\right), \tag{15}$$

$$x_{-1}, b_{-1}^\pi,\ b_{-1}^x, \gamma_0 \text{ given} \tag{16}$$

where the IS curve does not appear because it is never a binding constraint (the CB can always choose an interest rate that satisfies it, given the allocations and beliefs), and we used the NKPC to substitute out π_t.

Assuming that the CB influences beliefs is customary when private agents are rational, but it is less frequent when private agents are learning.[26] There is, however, a major difference between the two assumptions. Under RE, promises can influence beliefs. Under learning, the policymaker can influence beliefs exclusively through actions, i.e. by implementing output expansions and contractions (see (14) and (15)). With this assumption we address a common criticism of CB commitment, that it places too much faith on impacting private beliefs. We take the stance that it is important to understand the policy trade-offs at the other extreme, when only actions matter, because learning has been shown to be empirically relevant.[27] Undoubtedly, in practice, both promises and actions are important. In Section 5, we extend our analysis to a framework where both play a role.

[26] A few exceptions are Gaspar et al. [20] and Molnar and Santoro [31].

[27] There is no consensus yet on how to model learning, but several papers have shown its presence in private expectations. See, among others Branch and Evans [6], Milani [29], and Molnar and Ormeno [30].

The existence of a recursive solution[28] of the optimization problem (13) cannot be taken for granted, because of some nonstandard features: the updating rules for beliefs are not convex, the feasibility set is not compact-valued, and the quadratic return function is unbounded; however, in the Appendix we prove the following result:

Proposition 1. *There exists a time-invariant policy function for the CB that solves the optimization problem 13.*

Hence the solution to (13) can be characterized as the solution of the FOCs[29]:

$$\begin{aligned} 0 =& -\alpha x_t - \left[(\beta b^{\pi}_{t-1} + \kappa)x_t + u_t\right](\beta b^{\pi}_{t-1} + \kappa) - \lambda_{1,t}\gamma_t x_{t-1}(\beta b^{\pi}_{t-1} + \kappa) - \\ & - E_t[\lambda_{1,t+1}\beta\gamma_{t+1}((\beta b^{\pi}_t + \kappa)x_{t+1} + u_{t+1} - b^{\pi}_t 2x_t)] \end{aligned} \tag{17}$$

$$\begin{aligned} 0 =& \lambda_{1,t} - \beta E_t \lambda_{1,t+1}(1 - \gamma_{t+1}x_t^2) - \beta^2 E_t[((\beta b^{\pi}_t + \kappa)x_{t+1} + u_{t+1})\, x_{t+1}] - \\ & \beta^2 E_t\left[\lambda_{1,t+1}\gamma_{t+1}x_t x_{t+1}\right] \end{aligned} \tag{18}$$

where $\lambda_{1,t}$ is the Lagrange multiplier on (14).[30] These first-order conditions together with the law of motion for the learning coefficients constitute the necessary conditions for the optimal evolution of $\{x_t, b^{\pi}_t, b^{x}_t\}$.[31]

A key insight is that in the FOCs (17)-(18) all the terms that come from the manipulation of beliefs are weighted by the gain, and thus become irrelevant as $\gamma_t \to 0$, unless they grow unboundedly. In the Appendix we use this insight to rewrite the updating equations for beliefs as a stochastic recursive algorithm (SRA hereafter) in the standard form studied in Evans and Honkapohja [16]:

$$\theta_t = \theta_{t-1} + \gamma_t \mathcal{H}\left(\theta_{t-1}, Y_t\right) + \gamma_t^2 \rho\left(\theta_{t-1}, Y_t\right) \tag{19}$$

[28] Namely x_t, b^{π}_t, b^{x}_t as a time-invariant function of the five states $x_{t-1}, b^{\pi}_{t-1}, b^{x}_{t-1}, u_t, \gamma_t$; note that the learning dynamics implies that the parameters of beliefs (b^{π}, b^{x}) become natural state variables.

[29] We do not prove uniqueness of the optimal policy function, but it is not essential: in the analytical part we show asymptotic results valid for any optimal policy function, while in the numerical part we check that only one solution of the FOCs can be found.

[30] The Lagrange multiplier on (15) does not appear in the FOCs, because it can be shown that it is equal to 0 $\forall t$ in equilibrium; the proof is available upon request.

[31] From the IS curve and the NKPC we can back out the optimal processes for inflation and the nominal interest rate.

where $\theta_t \equiv [b_t^\pi, b_t^x]'$, $Y_t \equiv [x_t, x_{t-1}, u_t, \gamma_t,]'$, and all the terms coming from the manipulation of beliefs are grouped in the second-order term ρ.[32]

To study the asymptotic behavior of θ_t, we analyze the solutions and stability of the ordinary differential equation (ODE) associated to (19):

$$\frac{d\theta}{d\tau} = h(\theta) \equiv E\mathcal{H}(\theta, Y_t) \tag{20}$$

where the expectation is taken over the invariant distribution of the process $\widehat{Y}_t(\theta)$, which is the stochastic process for Y_t obtained by holding θ_{t-1} at the fixed value $\theta_{t-1} = \theta$.[33] Given the definition of $\mathcal{H}$ provided in the Appendix, we get:

$$h(\theta) = \begin{pmatrix} -b^\pi E x_{t-1}^2(\theta) \\ -b^x E x_{t-1}^2(\theta) \end{pmatrix}$$

The only possible rest point of the ODE (20) is clearly $\theta = 0$. Moreover it is (locally) stable, because the Jacobian:

$$Dh(\theta) = \begin{pmatrix} -Ex_{t-1}^2(\theta) - b^\pi \frac{\partial E x_{t-1}^2(\theta)}{\partial b^\pi} & -b^\pi \frac{\partial E x_{t-1}^2(\theta)}{\partial b^x} \\ -b^x \frac{\partial E x_{t-1}^2(\theta)}{\partial b^\pi} & -Ex_{t-1}^2(\theta) - b^x \frac{\partial E x_{t-1}^2(\theta)}{\partial b^x} \end{pmatrix} \tag{21}$$

has both eigenvalues smaller than zero when evaluated in $\theta = 0$.[34] In the terminology commonly used in the adaptive learning literature, we can say that $\theta = 0$ is the only *E-stable* equilibrium. From simple inspection of (21) we conclude that this E-stability result is independent of parameter values.

Remark 1. *The Jacobian* (21) *has negative eigenvalues for any value of the structural parameters.*

Evans and Honkapohja [16] derive an equivalence result between E-stability and convergence under learning. This theorem, which draws on arguments contained in

[32]For the exact definition of $\mathcal{H}$ and ρ, see the Appendix.

[33]It is possible to prove that there exists an invariant distribution to which the Markov process $\widehat{Y}_t(\theta)$ converges weakly from any initial conditions; hence, the function $h(\theta)$ is well defined. The proof is available from the authors upon request.

[34]We are implicitly assuming that $Ex_{t-1}^2(\theta)$ admits partial derivatives, and that they are finite.

Benveniste et al. [4], cannot directly be applied to our problem, because the state variables' law of motion does not satisfy the required assumptions.[35] However, we can prove the following result.[36]

Proposition 2. *Let θ evolve according to* (19). *If $\overline{\theta}$ is E-stable, then it is locally stable under adaptive learning.*[37]

Proposition 2 implies that in the limit $\theta_t = [b_t^\pi, b_t^x]' \to 0$. This is the only possible E-stable equilibrium and it is locally stable. Equation (10) then shows that in the limit agents expect zero inflation and output gap. Substituting this together with $\gamma_t \to 0$ into the FOC (17) and the PC (2) implies that both output and inflation converge to the IT equilibrium (6)-(7).

Main result 1. *The optimal policy drives the economy to the inflation targeting equilibrium*

$$x_t = -\frac{\kappa}{\alpha + \kappa^2} u_t, \quad \pi_t = \frac{\alpha}{\alpha + \kappa^2} u_t.$$

There are three striking features of our main result. First, it is optimal to implement an equilibrium that would be suboptimal under RE. In the limiting equilibrium, as private agents learn $b^x = b^\pi = 0$, the CB loses its ability to impact future allocations through current output contractions and expansions (see (8)-(9)), even though the CB would be able to retain this ability by implementing the PLT equilibrium. Second, although our result is valid only locally, our numerical simulations show that it holds irrespective of initial beliefs. No matter how close private beliefs are to the PLT equilibrium, even if initially the CB has "credibility" to implement PLT, it is optimal to drive the economy away from this equilibrium (for more on the role of

[35]From a technical point of view, the Markov chain followed by our state variables Y is not necessarily geometrically ergodic; hence, the assumption A.4 as stated on page 216 of Benveniste et al. [4] is not satisfied (we cannot prove the existence of a solution to the Poisson equation).

[36]Strictly speaking, the following result does not establish an equivalence between E-stability and convergence under learning, because it does not guarantee that any locally stable equilibrium is E-stable. However, our numerical investigation shows that this is the case.

[37]For an explicit definition of what "locally stable under adaptive learning" means, see Evans and Honkapohja [16] page 275.

"credibility", see Section 3).[38] Finally, our main result holds for any α in the welfare loss function. Even if the central banker cares strongly about dampening inflation fluctuations, i.e. α is low, it is optimal to deviate from PLT. Therefore the main result cannot be turned around by appointing a conservative central banker, in a way analogous to what was suggested in Rogoff [36].

3. Policy Implications

Policy incentives behind our main result are best illustrated by the unfolding dynamics. For presentational purposes, we will discuss simulations with constant gain learning, because it allows us to focus on the policy trade-offs while abstracting from the role of a changing gain parameter.[39] For our baseline simulations we set $\gamma = 0.05$, which is a value consistent with estimates for the US economy[40], and examine the role of the gain parameter at the end of Section 3.

3.1. Long- versus short-run policy trade-offs

Figure 1 illustrates our main result in welfare terms: as OP drives expectations asymptotically to the IT equilibrium, expected welfare losses increase to those of IT. For each time t, the figure plots the expected consumption equivalent (CE) measure of welfare losses (percentage of steady-state consumption) for an economy starting from period-t average beliefs; at time zero we start from PLT beliefs.[41] For comparison we plot the same CE measure for two Taylor-type rules, that Evans and Honkapohja [18] and Evans and Honkapohja [17] have proven to drive beliefs respectively to PLT

[38]In other words, imagine that a central banker inherits "credibility" from his predecessor in the sense that private expectations react to his policy as the PLT equilibrium prescribes. Our result then implies that, also in this case, there is an incentive to give up this ability.

[39]We simulate our economy with structural parameters of Woodford [43]: $\beta = 0.99, \sigma = 0.157, \kappa = 0.024, \alpha = 0.04, \sigma_u = 0.07$. Decreasing gain results are qualitatively similar to constant gain, but quantitatively sensitive to the exact timing. Results with decreasing gain are available upon request.

[40]See Milani [29] and Slobodyan and Wouters [40]

[41]We simulate 10,000 draws of 2000-period-long series, starting from beliefs corresponding to PLT at time 0, and we calculate the CE welfare loss. Then, we take the beliefs in period 1 for each one of the 10,000 draws, and from those beliefs we simulate 10,000 draws of 2000-period-long series, and then we calculate the CE welfare loss. We repeat this process for 8000 periods.

and IT equilibria. For the IT rule we set the initial beliefs at IT in order to illustrate the long-run welfare implications of keeping expectations in the IT equilibrium.[42]

The figure illustrates well why our main result is striking: the policymaker is fully rational and could induce the PLT equilibrium, which would be welfare enhancing in the long run, it is simply suboptimal to do so.

The *long-run* benefits of PLT would be similar to the case with rational agents, i.e. it anchors agents' inflation expectations once learning expectations have settled on the equilibrium; "keeping" learning expectations in the PLT equilibrium is superior to "keeping" them in the IT equilibrium. Similar results can be found also in different setups, which all show that expectations are better anchored under PLT. Preston [34] shows the robustness of long-term benefits of PLT to misinformation about agents learning[43]; in a framework featuring near-rational expectations, Woodford [45] argues that benefits of engineering a history-dependent policy are present also when expectations differ from RE with a nonspecified error structure.

However, it is optimal to sacrifice long-run efficiency for *short-run gains*. By starting from PLT beliefs we are implicitly assuming that initially the CB has "credibility", i.e. it can reduce inflation expectations by contracting output. It is in these initial periods that our optimal policy can generate lower welfare losses than PLT, because it can exploit the sluggish nature of expectations. While PLT anchors future inflation expectations by committing to spread out the effect of shocks, OP can respond more aggressively to shocks because the policymakers' credibility will not be harmed in the short-run. Agents need to gather enough data to uncover a deviation from the PLT. Even if credibility is lost in the long run, short-run gains far outweigh long-run losses: expected CE of PLT is 63% higher than that of OP when agents

[42] The main appeal of these rules is that besides ensuring stability under learning, they also guarantee determinacy under RE. A caveat shown in Preston [33] is that under infinite horizon learning, these rules can induce divergent learning dynamics, because the CB does not give enough attention to future private expectations.

[43] Preston [34] examines one-period-ahead expectations-based Taylor rules, whereas agents have infinite horizon learning. We will relax the assumption of perfect knowledge of agents' learning in Section 5.

initially believe in a PLT policy (see Table 1).[44]

Even though the CB takes advantage of its credibility during the transition, it has *no incentive to build credibility* at any point in time. As the CB keeps engineering surprise output contractions, expectations keep getting further away from PLT, and agents believe less and less in a history-dependent policy (see Figure 2). OP is however careful *not to lose credibility too fast*, in order to maintain its ability to disinflate through lowering inflation expectations (i.e. keep $b^\pi > 0$, such that $\hat{E}\pi_{t+1} = b^\pi x_t$ can be lowered by lowering x_t). Based on forecast errors, it would not be easy for agents to conclude that the CB deviated from PLT (for more on this, see Section 5). First, they are small during the transition, similar in size to what would arise in the PLT equilibrium (Figure 3).[45] Second, there is no systematic pattern in forecast errors: agents sometimes overpredict, sometimes underpredict the outcome (see Figure 4). Only when the economy converges close enough to IT do forecast errors increase, as the CB loses its incentive to keep inflation expectations history dependent. Where the CB really fools agents is in output expectations, but these have a small impact on welfare losses.[46] As the economy converges on IT, forecast errors become similar to those of a rational agent in IT. All these forecast errors are however very small in magnitude.

The way CB credibility is lost is fundamentally different for learning and rational agents. Any deviation from a commitment is immediately spotted by rational agents, making any future commitment of the CB not credible anymore. This off-equilibrium threat helps maintain the PLT equilibrium. Learners lack off-equilibrium strategies, as they learn only from realized outcomes, and during this learning process the policymaker has an incentive to deviate from PLT. Speeding up learning does not eliminate these CB incentives, it merely reduces them. We can see this in Figure

[44] Note, that in our setup PLT and IT consumption equivalents are both small, albeit in the range of the original estimates of Lucas [27].

[45] A rational agent in the PLT equilibrium would have an expected squared forecast error of $c^{PLT}\sigma_u^2 = 0.0039$.

[46] For a bigger weight of output in the welfare loss function, α, forecast errors of output decrease, and of inflation increase.

5: for a bigger γ OP engineers less-aggressive output contractions in response to a positive cost-push shock.[47]

The loss of credibility in the long run cannot be solved by delegation, in the spirit of Rogoff [36], by appointing a more patient central banker (higher β).[48] As long as future losses are discounted, $\beta < 1$, in the long run IT is the resulting equilibrium. We can observe in Figure 7b that all a more patient central banker achieves is keeping the economy close to the welfare-improving PLT equilibrium for a longer period, i.e. retaining "credibility" longer, because she is exploiting less the short-run policy trade-offs.[49]

Table 1: Consumption equivalents

	OP	PLT	ratio PLT/OP
Initial beliefs:			
PLT	0.000413	0.000675	1.63
IT	0.000747	0.001004	1.34

3.2. Short-run policy incentives

The short-run gains of OP come from the well-known time-inconsistency problem of PLT and the sluggishness of agents' beliefs. The time inconsistency is standard: if given the chance, the CB has an incentive to renege its commitments and choose a different policy that is optimal at the time the decision is taken.

This incentive to deviate from PLT can be easily illustrated in a simple case, when agents do not update their learning coefficients ($\gamma_t = 0$). The joint FOCs do not depend on x_{t-1}, as in the PLT equilibrium; instead the strategy is similar to that

[47] In Section 5 we return to examine whether these CB incentives would survive with other expectation formations.

[48] In contrast to the original Rogoff [36] problem, where delegation aims to solve the inflation bias, here we think of a delegation that aims to solve the bias for short-term gains.

[49] A higher resemblance to credibility with higher patience is also shown in Sargent [39] and Molnar and Santoro [31], who also analyze learning environments. Sargent [39], Chapter 5, obtains the remarkable result that the optimal policy in the Phelps problem is such that a CB which is patient enough ($\beta \to 1$) can replicate the commitment solution under RE asymptotically. Eusepi et al. [12] obtain similar results in a New Keynesian model investigating the optimal long-run inflation rate, rather than dynamic responses to shocks, as we do in this paper.

of the "leaning against the wind" of IT: after a positive shock, the CB decreases the current output gap in order to avoid a huge increase in current inflation.

$$\begin{aligned} x_t &= -\frac{\beta b^\pi + \kappa}{\alpha + (\beta b^\pi + \kappa)^2} u_t \\ \pi_t &= \frac{\alpha}{\alpha + (\beta b^\pi + \kappa)^2} u_t. \end{aligned} \tag{22}$$

The output contraction is stronger the more credible the CB is: the higher is b^π, the stronger is the trade-off between inflation and output (from (2)), and therefore the stronger is the incentive of the CB to "fool" agents.

Similar incentives arise when agents are learning, because learning takes time. Agents need to collect sufficient data to understand if the CB deviates from PLT. As in the case with $\gamma = 0$, the further beliefs are from the IT equilibrium, the larger is the surprise output contraction engineered by the CB, because the larger is the policy incentive to exploit the inflation-output trade-off (Figure 5).

As OP aims to lower inflation, it lets prices absorb shocks in a permanent way: after a positive cost-push shock the price level raises permanently (see Figure 6c). This is similar to an IT rule, which would treat a cost-push shock as bygone. In contrast, under PLT the CB would bring the price level back to the target.

The main difference between our policy and previously proposed Taylor rules, is that our policy is *nonlinear in agents' beliefs*. (see Figure 5). OP exploits the fact that the closer households' beliefs are to the PLT equilibrium, the larger is the output contraction that can be engineered without loss of "credibility". In contrast, the Taylor rule that implements PLT is a linear: the further away beliefs are from the PLT equilibrium, the larger the output contraction that PLT policy engineers, in order to drive beliefs back to the PLT equilibrium.

4. Implementation with a simple rule

We now turn to the question of how policy should be conducted. Deriving an analytical policy rule for the optimal state-contingent interest rate path is a nontrivial task, because it is a highly nonlinear rule in agents' beliefs and their speed of learning.

This nonlinearity would also make its implementation impractical. Moreover, such a rule would require detailed knowledge of how agents learn. Such superior knowledge on the part of the policymaker is a very strong assumption (see Woodford [45]).

However, it turns out the OP can be well approximated without knowing the exact form of agents' learning. Consider the following simple belief-dependent Taylor-type policy rule, which is obtained by solving problem (13) when $\gamma = 0$.

$$i_t = \delta_\pi(b^\pi_{t-1})E^*_t\pi_{t+1} + \delta_x E^*_t x_{t+1} + \delta_u(b^\pi_{t-1})u_t, \tag{23}$$

where the coefficients are

$$\delta_\pi(b^\pi_{t-1}) = 1 + \sigma\beta\frac{\beta b^\pi_{t-1} + \kappa}{\alpha + \kappa^2 + \kappa\beta b^\pi_{t-1}}, \quad \delta_x = \sigma, \quad \delta_u(b^\pi_{t-1}) = \sigma\frac{\beta b^\pi_{t-1} + \kappa}{\alpha + \kappa^2 + \kappa\beta^\pi_{t-1}} \tag{24}$$

This rule satisfies the Taylor principle, and guarantees both determinacy under RE and E-stability[50], hence yielding convergence under learning.

A desirable property of this rule is that it achieves most of the welfare gain that OP achieves: its CE is less than 5% higher than that of OP, when initial beliefs are at PLT.[51] Moreover, it is simple: the policymaker only needs to monitor inflation and output gap expectations at each point in time, without the need to gauge the effect on the evolution of future expectations. To determine the coefficients of this rule, the policymaker also needs to know that agents understand monetary policy-making to some extent, so that an output contraction lowers inflation expectations. Central banks do indeed dedicate substantial time and effort to both: they monitor expectations and also educate the general public about the conduct of monetary policymaking.[52].

The main difference with respect to the expectations-based rule suggested in Evans and Honkapohja [14] is that our policymaker also exploits the knowledge that agents understand policymaking to some extent, i.e. they reduce inflation expecta-

[50]The eigenvalues of the reduced form are 0 and $0 < \frac{\beta\alpha}{\alpha+\kappa^2+\kappa\beta b^\pi_{t-1}} < 1$, for $b^\pi_{IT} < b^\pi_{t-1} < b^\pi_{PLT}$.

[51]Details available upon request.

[52] Carvalho and Nechio [7] show evidence from survey expectations that people do understand monetary policy: this educational effort seems to work.

tions when output contracts. As a consequence, our simple rule contracts output somewhat more aggressively to bring down inflation expectations (higher δ_π).

5. Extensions

A few of our assumptions play an important role in our findings, and thus we would like to discuss and examine their limitations.[53]

5.1. Generalizability

We conduct our analysis by assuming a specific learning algorithm. This algorithm can be justified on several grounds[54] but it might seem arbitrary because it is just one out of many. Yet, it is equally arbitrary to assume that there is no learning component to private expectations, especially because this is at odds with recent empirical findings.[55] Ultimately, how people form expectations is a yet unsettled issue, which presents an important challenge for policymakers. Would they know the exact expectation formation of agents, policy would be much easier. Then emerges the question on what are the limits of our results, and whether they are robust to the existence of different classes of beliefs.

We expand our analysis to a general class of learning algorithms, where belief updating is a general function of past output gaps, forecast error and the cost-push shock.[56]We build beliefs updating on Selten's *directional learning*: we simply assume that learners have enough knowledge to determine myopically in which direction

[53]We refer the reader to the Appendix for details.

[54]It is widely used in the literature, consistent with the rational expectation equilibrium, and empirically relevant.

[55]See for example [5] on evidence from survey expectations and [9] on experiments.

[56]M^π, M^x twice continuously differentiable, equal to zero if and only if the forecast errors are equal to zero, and increasing in the forecast error if and only if $x_{t-1} > 0$: if agents expect a positive π_t, i.e. $b^\pi_{t-1}x_{t-1}$ is positive, and π_t turns out to be even more positive, agents want to increase $b^\pi_t x_t$ to track π.

better forecasts can be found.[57]

$$b_t^\pi = b_{t-1}^\pi + \gamma_t M^\pi \left(x_{t-1}, \pi_t - b_{t-1}^\pi x_{t-1}, u_t\right) \tag{25}$$
$$b_t^x = b_{t-1}^x + \gamma_t M^x \left(x_{t-1}, x_t - b_{t-1}^x x_{t-1}, u_t\right),$$

This general formulation includes, as special cases, the stochastic gradient we used in the baseline specification and the generalized stochastic gradient introduced in Evans et al. [19].

We find that, consistent with our main result, learning PLT is never optimal when agents use (25) to update beliefs. As in the baseline analysis (see Section 3.2), the incentives for the CB to deviate from PLT are related to the result that in the limit for $t \to \infty$ the *trade-off between inflation and output gap is not affected by learning*; in other words, CB cannot manipulate beliefs anymore, and it pursues a "lean against the wind" policy: whenever inflation is high, contract demand below capacity (and vice versa).[58]

$$\pi_t = -\frac{\alpha}{\beta b^\pi + \kappa} x_t \tag{26}$$

This result is therefore not specific to the form of learning we adopted in the baseline analysis, but is more general: what matters is that *off-equilibrium*, when the CB deviates from PLT, agents change their beliefs in a gradual and adaptive manner, and not abruptly as a rational agent would do. However, the linearity of M in the forecasting error is necessary for IT to be an equilibrium. In this case $b^\pi = b^x = 0$ is a solution to $h(\theta) = 0$. For a nonlinear updating this is not necessarily the case, which implies that, theoretically, nonlinear learning can converge to an equilibrium different from IT.

Next, we examine robustness to the *presence of rational agents* next to learners.

[57] An alternative interpretation of directional learning which works even if subjects have very little information is trial-and-error learning. It simply says that an agent would not repeat a mistake, i.e. if forecasts last period have overestimated the outcome then one would not increase forecasts again.

[58] Mathematically this result is a consequence of the following: (under suitable technical conditions) all the terms which come from beliefs' manipulation are weighted by the learning gain. In the limit they become irrelevant as t goes to zero, unless they grow unboundedly (proof in Appendix).

If the CB has commitment with respect to the rational agents (who have a $(1-\psi)$ population weight), the optimality condition in the limit is similar to (26), with an additional path dependent term that is introduced as a consequence of the promises made by the central bank and trusted by the rational agents

$$\pi_t = -\frac{\alpha}{\beta\psi b^{\pi} + \kappa} x_t + (1-\psi)\frac{\alpha}{\beta\psi b^{\pi} + \kappa} x_{t-1}. \tag{27}$$

In the long run the economy converges somewhere between IT and PLT: optimal allocations have some history dependence. This means the CB can retain some credibility, but quantitatively the impact is very small even when there are many rationals in the population. When half of the population is rational, for example, in the limiting equilibrium the b^{π} is an order of magnitude smaller than in PLT. Thus deviation from PLT is a robust result, unless all agents are rational.

5.2. Uncertainty about learning, and evolutionary dynamics

Our main result also relies on the assumption that policymakers perfectly understand agents' belief formation. This assumption is routinely made under RE but is less innocuous under learning: there is one way to be rational, but infinite ways to be nonrational. To examine robustness, we hypothesize a CB that can face several empirically relevant learning algorithms,[59] and find that using our baseline OP rule outperforms PLT. For policymaking, it is more important to know agents learn than to gauge how exactly they do it.

Finally, in our main analysis we presumed little thinking on the agents side, while the policymaker is strategic.[60] This raises the question, whether agents would leave their expectation formation if they could. We endow agents with such "evolutionary"

[59] We assume OP with our baseline learning specification ($\gamma = 0.05$) whereas agents learning is different (they can have a different γ, or learn with a decreasing gain).

[60] See Woodford [45] who cautions about strategic manipulation by the policymaker of agents' learning rules: "... the CB can induce systematic forecasting errors of a kind that happen to serve the central bank's stabilization objectives. But if such a policy were shown to be possible under some model of learning considered to be plausible (or even consistent with historical data), would it really make sense to conduct policy accordingly, relying on the public to continue making precisely the mistakes that the policy is designed to exploit?"

skills and find that even if they can switch towards a fully rational expectation formation (if it forecasts better) in the limiting equilibrium learning survives. The reason for this is that the learning mechanism produces good forecasts compared to RE: initially the policymaker keeps learners' forecast errors small (similar to Section 3), and in the limit learners learn to forecast as well as rational agents.

6. Concluding remarks

We have argued that the benefits of PLT hinges not only on a skillful management of expectations but also on agents being rational. If we relax rationality bounds on agent's understanding, stabilizing prices is a bad strategy. In the context of adaptively learning agents we contend that monetary policy has strong short-run incentives to deviate from PLT, despite its benefits in effectively anchoring inflation expectations. These incentives arise because learning agents need time to discover that the CB has deviated from PLT, and in the meantime the policymaker can exploit the inflation-output trade-off and disinflate by aggressively contracting output. This policy comes at a cost: private agents eventually gather enough data and understand that the CB is deviating from PLT. The economy converges on IT and the CB loses its ability to anchor private expectations. We show that the short-run gains of this policy outweigh long-run losses, and therefore it is optimal for the CB to succumb to the temptation and deviate from PLT.

In our main analysis we assume the CB knows the exact learning algorithm, which is a strong assumption.[61] Therefore we have also established that for policymaking, the most important welfare gains arise from knowing that agents learn, and it is of second order to gauge how exactly agents update beliefs. Finally, we have shown generalizability of short-run incentives to deviate from PLT for a general learning algorithm, and for a hybrid model, with intermediate forms of rationality mixing

[61] Note that an analogously strong assumption is regularly made in optimal policy research with rational agents, where the policymaker knows that agents are rational. We think it is worth making our extreme assumption in order to understand optimal policy under the polar case of adaptive learning, given the empirical relevance of learning in survey and experimental evidence (see for example Del Negro and Eusepi [10], Slobodyan and Wouters [40], Molnar and Ormeno [30]).

rational and adaptive agents.

The CB incentives that arise in our framework have previously been ignored by proponents of PLT under learning (see Evans and Honkapohja [18], Aoki and Nikolov [2], Gaspar et al. [22]). Those authors showed that PLT is a learnable equilibrium: if expectations are perturbed out of the PLT equilibrium, the CB can implement a policy that makes agents learn the PLT equilibrium again. However, once CB incentives are taken into account, PLT is no longer optimal if agents are learning.

A general message from our results is that in a heterogenous agents setup, it is not enough to examine the learnability of an equilibrium, as it is traditionally done in the literature (see Evans and Honkapohja [16]). Even a learnable equilibrium might not arise when interactions between agents are taken into account. The incentives of a rational player (in our model the CB) depend on what type of other player she interacts with. Adaptive players are different from rational players even after they learned a rational expectations equilibrium, and their forecasts could not be distinguished from those of a rational agent. One difference is the speed of revising beliefs. A rational agent would immediately understand if the CB has deviated from PLT and would immediately switch to the IT equilibrium. A learning agent on the other hand needs time to gather a sufficient amount of data to understand that the CB deviated from PLT. A second, more subtle difference is that rational agents can choose a strategy that prescribes totally different behavior on- and off-equilibrium, and the off-equilibrium threat of rational private agents can keep a rational bank from deviating from PLT (see Kurozumi [26]). For learning agents, on the other hand, off-equilibrium threats are not possible, because they simply form beliefs based on realized outcomes. A rational opponent to learning agents takes this into account and chooses her strategy accordingly.

Finally, let us note that we do not mean to give precise policy prescriptions to central banks. We are aware that policymaking in reality is more complex and challenging than in our simple framework. Our results however should highlight that the incentives of the CB change with the belief structure of the private sector, and policy prescriptions derived without acknowledging this fact can be misleading.

References

[1] Adam, K., Marcet, A., 2011. Internal rationality, imperfect market knowledge and asset prices. Journal of Economic Theory 146, 1224–1252.

[2] Aoki, K., Nikolov, K., 2006. Rule-Based Monetary Policy under Central Bank Learning, in: NBER International Seminar on Macroeconomics 2004. National Bureau of Economic Research, Inc. NBER Chapters, pp. 145–195.

[3] Barucci, E., Landi, L., 1997. Least mean squares learning in self-referential linear stochastic models. Economics Letters 57, 313 – 317.

[4] Benveniste, A., Métivier, M., Priouret, P., 1990. Adaptive Algorithms and Stochastic Approximations. Berlin: Springer-Verlag.

[5] Branch, W.A., 2004. The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations. The Economic Journal 114, 592–621.

[6] Branch, W.A., Evans, G.W., 2006. A simple recursive forecasting model. Economics Letters 91, 158–166.

[7] Carvalho, C., Nechio, F., 2014. Do people understand monetary policy? Journal of Monetary Economics 66, 108 – 123.

[8] Clarida, R., Gali, J., Gertler, M., 1999. The science of monetary policy: A new keynesian perspective. Journal of Economic Literature 37, 1661–1707.

[9] Dal Bo, P., Frechette, G.R., 2011. The evolution of cooperation in infinitely repeated games: Experimental evidence. American Economic Review 101, 411–29.

[10] Del Negro, M., Eusepi, S., 2011. Fitting observed inflation expectations. Journal of Economic Dynamics and Control 35, 2105–2131.

[11] Eggertsson, G.B., Woodford, M., 2003. The Zero Bound on Interest Rates and Optimal Monetary Policy. Brookings Papers on Economic Activity 34, 139–235.

[12] Eusepi, S., Giannoni, M.P., Preston, B., 2018. Some implications of learning for price stability. European Economic Review 106, 1 – 20.

[13] Eusepi, S., Preston, B., 2016. The science of monetary policy: an imperfect knowledge perspective. Staff Reports. Federal Reserve Bank of New York.

[14] Evans, G.W., Honkapohja, S., 1998a. Convergence of learning algorithms without a projection facility. Journal of Mathematical Economics , 59–86.

[15] Evans, G.W., Honkapohja, S., 1998b. Stochastic gradient learning in the cobweb model. Economics Letters 61, 333 – 337.

[16] Evans, G.W., Honkapohja, S., 2001. Learning and Expectations in Macroeconomics. Princeton: Princeton University Press.

[17] Evans, G.W., Honkapohja, S., 2003. Expectations and the stability problem for optimal monetary policies. Review of Economic Studies 70, 807–824.

[18] Evans, G.W., Honkapohja, S., 2006. Monetary policy, expectations and commitment. The Scandinavian Journal of Economics 108, 15–38.

[19] Evans, G.W., Honkapohja, S., Williams, N., 2010. Generalized stochastic gradient learning. International Economic Review 51, 237–262.

[20] Gaspar, V., Smets, F., Vestin, D., 2006a. Adaptive Learning, Persistence, and Optimal Monetary Policy. Journal of the European Economic Association 4, 376–385.

[21] Gaspar, V., Smets, F., Vestin, D., 2006b. Optimal Monetary Policy under Adaptive Learning. Computing in Economics and Finance 2006 183. Society for Computational Economics.

[22] Gaspar, V., Smets, F., Vestin, D., 2007. Is time ripe for price level path stability? Working Paper Series 0818. European Central Bank.

[23] Gorodnichenko, Y., Shapiro, M.D., 2007. Monetary policy when potential output is uncertain: Understanding the growth gamble of the 1990s. Journal of Monetary Economics 54, 1132–1162.

[24] Hatcher, M., Minford, P., 2016. Stabilisation policy, rational expectations and price-level versus inflation targeting: a survey. Journal of Economic Surveys 30, 327–355.

[25] Honkapohja, S., Mitra, K., Evans, G.W., 2011. Notes on Agents' Behavioral Rules Under Adaptive Learning and Studies of Monetary Policy. SIRE Discussion Papers 2011-04. Scottish Institute for Research in Economics (SIRE).

[26] Kurozumi, T., 2008. Optimal sustainable monetary policy. Journal of Monetary Economics 55, 1277–1289.

[27] Lucas, R.E., 1987. Models of Business Cycles. Oxford: Basil Blackwell.

[28] Milani, F., 2006. A bayesian dsge model with infinite-horizon learning: Do "mechanical" sources of persistence become superfluous? International Journal of Central Banking 2, 87–106.

[29] Milani, F., 2007. Expectations, learning and macroeconomic persistence. Journal of Monetary Economics 54, 2065–2082.

[30] Molnar, K., Ormeno, A., 2015. Using survey data of inflation expectations in the estimation of learning and rational expectations models. Journal of Money Credit and Banking 47, 673–699.

[31] Molnar, K., Santoro, S., 2014. Optimal monetary policy when agents are learning. European Economic review 66, 39–62.

[32] Preston, B., 2005. Learning about Monetary Policy Rules when Long-Horizon Expectations Matter. International Journal of Central Banking 1.

[33] Preston, B., 2006. Adaptive learning, forecast-based instrument rules and monetary policy. Journal of Monetary Economics 53, 507–535.

[34] Preston, B., 2008. Adaptive learning and the use of forecasts in monetary policy. Journal of Economic Dynamics and Control 32, 3661 – 3681.

[35] Reis, R., 2013. Central bank design. Journal of Economic Perspectives 27, 17–44.

[36] Rogoff, K., 1985. The Optimal Degree of Commitment to an Intermediate Monetary Target. The Quarterly Journal of Economics 100, 1169–89.

[37] Røisland, Ø., 2006. Inflation inertia and the optimal hybrid inflation/price-level target. Journal of Money, Credit and Banking 38, 2247–2251.

[38] Rotemberg, J.J., Woodford, M., 1997. An optimization-based econometric framework for the evaluation of monetary policy, in: Bernanke, B., Rotemberg, J.J. (Eds.), NBER Macroeconomics Annual 12. Cambridge (MA): MIT Press.

[39] Sargent, T.J., 1999. The Conquest of American Inflation. Princeton: Princeton University Press.

[40] Slobodyan, S., Wouters, R., 2012. Learning in an estimated medium-scale dsge model. Journal of Economic Dynamics and Control 36, 26–46.

[41] Vestin, D., 2006. Price-level versus inflation targeting. Journal of Monetary Economics 53, 1361–1376.

[42] Wolman, A.L., 2005. Real Implications of the Zero Bound on Nominal Interest Rates. Journal of Money, Credit and Banking 37, 273–96.

[43] Woodford, M., 1999. Optimal monetary policy inertia. The Manchester School, Supplement 67, 1–35.

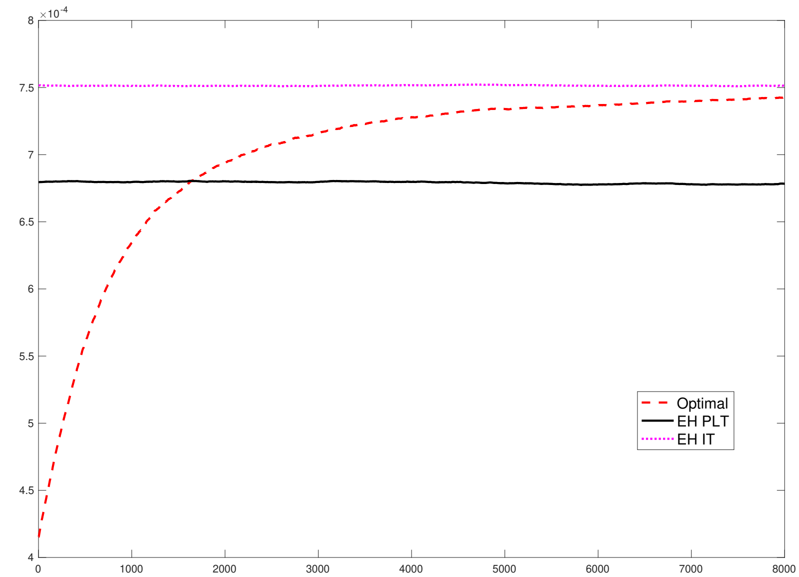
[44] Woodford, M., 2003. Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton: Princeton University Press.

[45] Woodford, M., 2010. Robustly Optimal Monetary Policy with Near-Rational Expectations. American Economic Review 100, 274–303.

[46] Yun, T., 1996. Nominal price rigidity, money supply endogeneity, and business cycles. Journal of Monetary Economics 37, 345–370.

7. Figures and Tables

Figure 1: Consumption equivalents losses, on a rolling window



Montecarlo of 10000 simulations. Initial beliefs at price level targeting for OP and PLT, at inflation targeting for IT, $\gamma = 0.05$.

Figure 2: Evolution of learning coefficient over time for different initial beliefs, ranging from IT to PLT

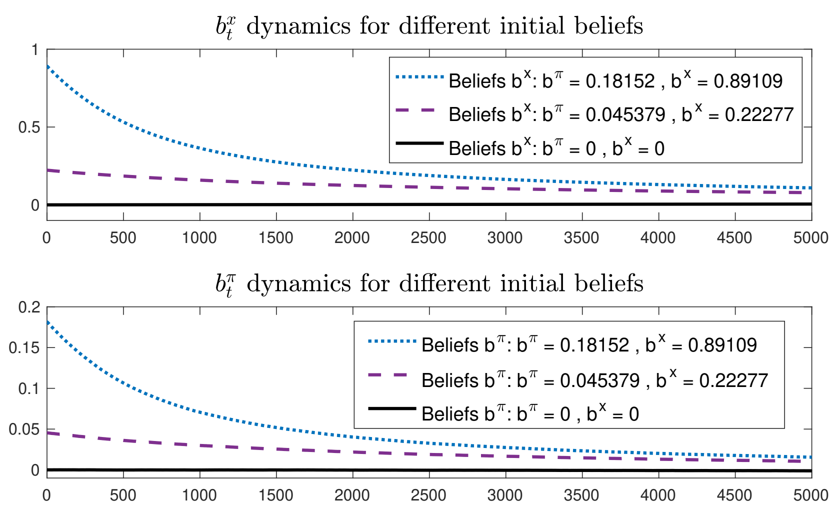
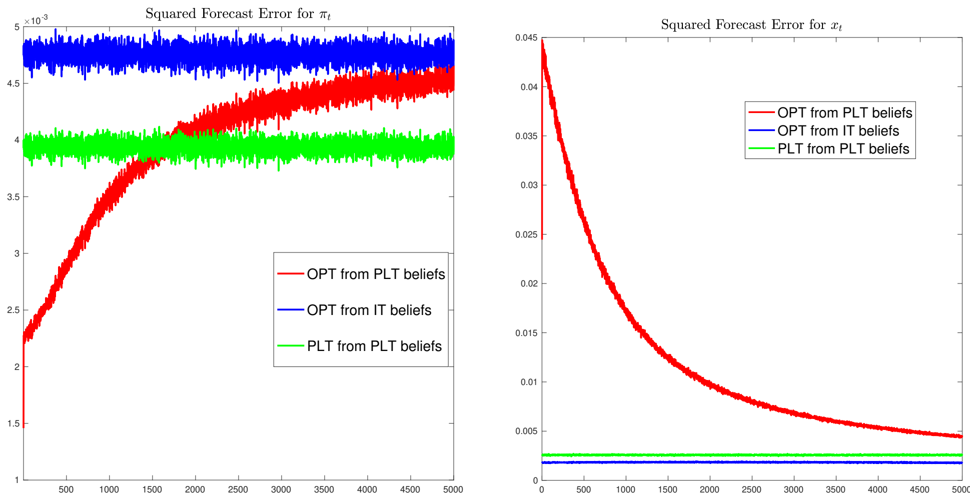


Figure 3: Squared forecast errors



(a) Squared forecast errors for inflation, for OP starting from different initial conditions and for PLT

(b) Squared forecast errors for output gap, for OP starting from different initial conditions and for PLT

Figure 4: Forecast errors for output gap and inflation for one series, OP starting from PLT

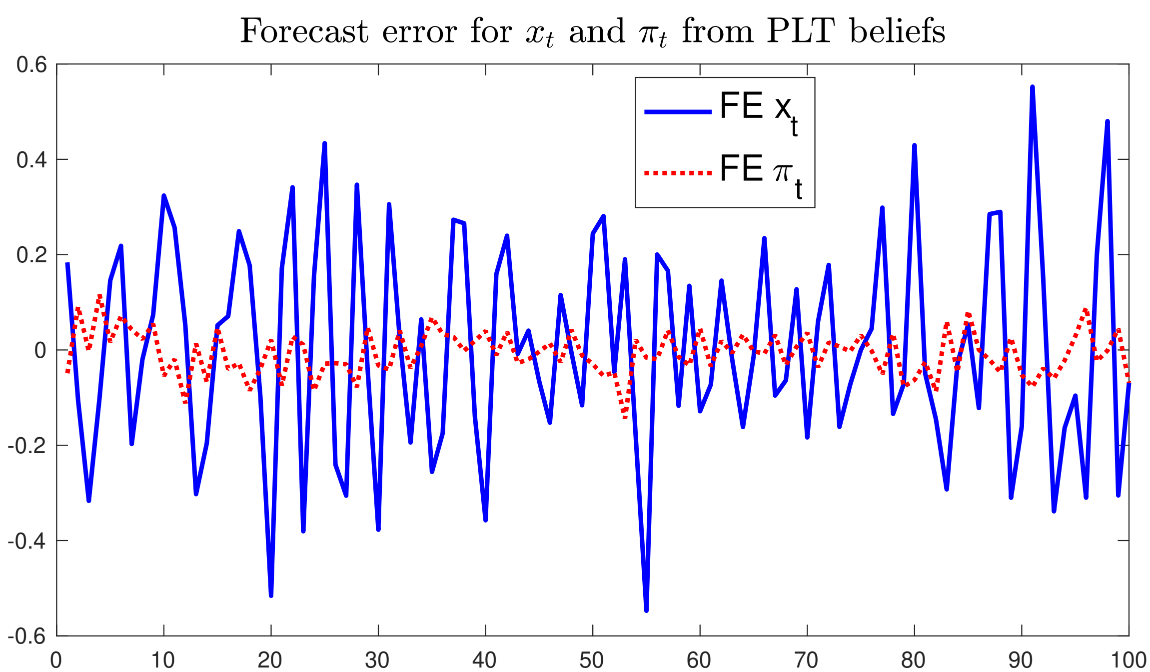
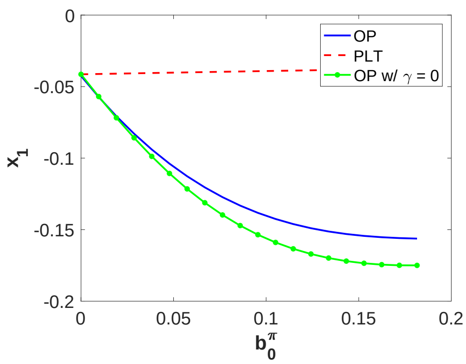
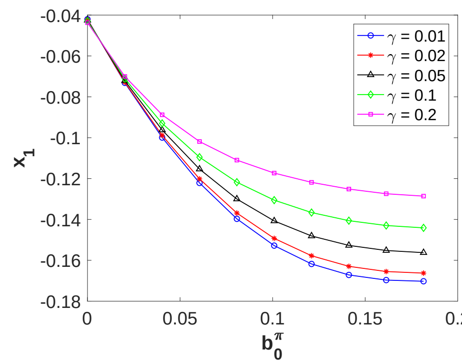


Figure 5: On-impact output gap responses with different private sector beliefs (to a one standard deviation cost-push shock)

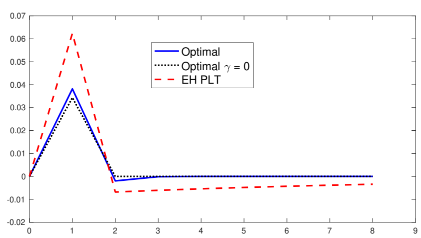


(a) Impulse effect on output gap as a function of beliefs, for OP and PLT policies

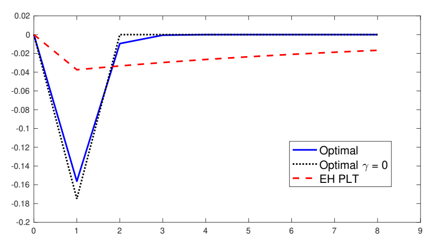


(b) Impulse effect on output gap as a function of beliefs, for different γ tracking parameters

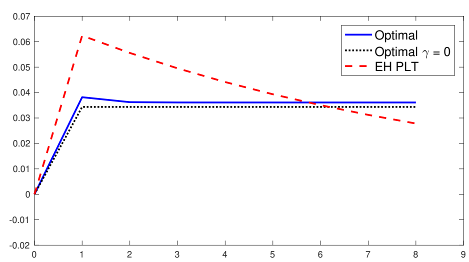
Figure 6: Impulse responses after a one standard deviation cost-push shock, under optimal policy under learning (OP) and price-level targeting policy (PLT) , starting with initial beliefs corresponding to the rational expectations PLT equilibrium, with $\gamma = 0.05$.



(a) Inflation π_t

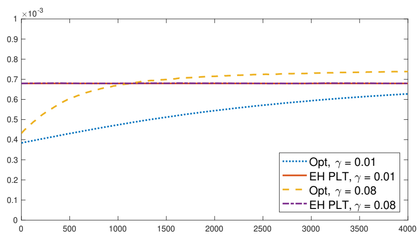


(b) Output-gap x_t

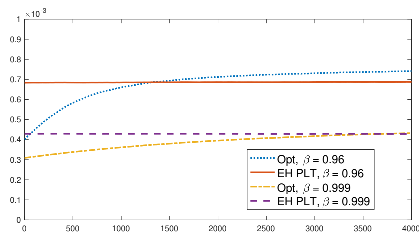


(c) Price level p_t

Figure 7: Consumption equivalents



(a) Slow and fast learning (low and high γ)



(b) Patient and impatient CB (high and low β)