

1 On the perils of stabilizing prices when agents are learning

2 Antonio Mele¹, Krisztina Molnár² and Sergio Santoro³

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. Disinflation 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.

4 JEL classification: C62, D83, D84, E52

5 No monetary authority sets price level stabilization⁴ as its official goal, despite
6 a vast literature claiming that it is a serious contender as a good way to conduct

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

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

³European 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.

⁴Price 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

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

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

24 In our setup, there is a stabilization role for monetary policy, i.e. reducing eco-
25 nomic fluctuations by dampening the effect of shocks on aggregate variables. Firms
26 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.)

⁵In 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].

⁶Swedish 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.

⁷This 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.”

27 how aggregate allocations are impacted by monetary policy. If their understanding
28 were perfect, they could form accurate expectations about how equilibrium alloca-
29 tions depend on shocks. This is the standard rational expectations assumption, and
30 in this case it is a well-established result (see for example Clarida et al. [8] and Vestin
31 [41]) that it is optimal to stabilize prices. The advantage of this policy arises from
32 its history dependence: after a temporary shock that increases the price level, the
33 policymaker engineers a series of aggregate demand contractions in order to bring
34 the price level back to its target; in other words, it can spread out the effect of the
35 shock on the price level through several periods. If agents are aware of this history
36 dependence, the policymaker can lower agents' expectations about future inflation
37 by contracting current output. Lower inflation expectations then decrease current
38 inflation through the Phillips Curve.⁸

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

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

⁸Our 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.

⁹We 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.

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

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

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

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

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

¹⁰Eusepi 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.

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

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

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

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

109 The policymaker has no incentive to build credibility (in the sense that it can

110 anchor inflation expectations by contracting output). Along the transition, as long
111 as the CB has some credibility, it also has an incentive to exploit it. In the long run,
112 when agents learn to ignore output contractions in forming their inflation expecta-
113 tions, temporarily revamping even little credibility becomes too costly for the CB,
114 especially because it would lose it immediately.

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

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

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

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

¹²PLT 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].

136 **1. The Model**

137 We develop our idea by weakening the assumption on private sector expectations
 138 in the well-known monetary policy analysis of Clarida et al. [8]. This example is
 139 chosen because the policy implications under rational expectations are well-known
 140 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

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

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

141 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^*
 142 denotes conditional expectations of the private sector, which are not necessarily ra-
 143 tional. The analysis is simplified by assuming that agents have common expectations,
 144 and have common knowledge about this; given this the linear aggregate relations can
 145 be derived with the usual log-linear approximation to equilibrium relations.¹⁵

146 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].

¹⁴This 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.

¹⁵As 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.

¹⁶The 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.

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

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

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

160 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 motion¹⁸:

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

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

161 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 $b^\pi = \frac{\alpha}{\kappa}(1 - b^x)$, $c^\pi = -\frac{\alpha}{\kappa}c^x$.
 162 This policy is equivalent to PLT: the CB responds to changes in the price level,
 163 and tries to keep prices close to a predetermined value. In equilibrium the price

¹⁷Because 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.

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

164 level follows a stationary process.¹⁹ The advantage of price level stabilisation arises
 165 from its history dependence: in a forward-looking environment history dependence
 166 entails welfare gains, because the policymaker can lower agents' expectations about
 167 future inflation by contracting current output and spreading the cost of adjustment
 168 to shocks over several periods. This history dependence is a robust feature of the
 169 optimal policy, even in setups more complicated than ours (see Hatcher and Minford
 170 [24]).

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

$$x_t = -\frac{\kappa}{\alpha + \kappa^2}u_t \quad (6)$$

$$\pi_t = \frac{\alpha}{\alpha + \kappa^2}u_t \quad (7)$$

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

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

177 1.2. Learning specification

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

¹⁹The 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 $\gamma \equiv \alpha/(\alpha(1 + \beta) + \kappa^2)$.

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

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

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

$$\pi_t = b^\pi x_{t-1} + c^\pi u_t \quad (8)$$

$$x_t = b^x x_{t-1} + c^x u_t, \quad (9)$$

192 Under learning, agents estimate the coefficients in equations (8)-(9), and use their
 193 estimates of b_{t-1}^π and b_{t-1}^x and the i.i.d. nature of u_t to make forecasts²²:

$$E_t^* \pi_{t+1} = b_{t-1}^\pi x_t, \quad E_t^* x_{t+1} = b_{t-1}^x x_t \quad (10)$$

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

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

²¹Agents 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.

²²Agents 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.

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

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

$$b_t^\pi = b_{t-1}^\pi + \gamma_t x_{t-1} (\pi_t - x_{t-1} b_{t-1}^\pi) \quad (11)$$

$$b_t^x = b_{t-1}^x + \gamma_t x_{t-1} (x_t - x_{t-1} b_{t-1}^x), \quad (12)$$

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

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

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

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

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

223 **2. Optimal monetary policy**

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

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

s.t.

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

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

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

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

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

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

²⁷There 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].

241 The existence of a recursive solution²⁸ of the optimization problem (13) cannot
 242 be taken for granted, because of some nonstandard features: the updating rules for
 243 beliefs are not convex, the feasibility set is not compact-valued, and the quadratic
 244 return function is unbounded; however, in the Appendix we prove the following
 245 result:

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

Hence the solution to (13) can be characterized as the solution of the FOCs²⁹:

$$0 = -\alpha x_t - [(\beta b_{t-1}^\pi + \kappa)x_t + u_t] (\beta b_{t-1}^\pi + \kappa) - \lambda_{1,t} \gamma_t x_{t-1} (\beta b_{t-1}^\pi + \kappa) - \quad (17)$$

$$- E_t[\lambda_{1,t+1} \beta \gamma_{t+1} ((\beta b_t^\pi + \kappa)x_{t+1} + u_{t+1} - b_t^\pi 2x_t)]$$

$$0 = \lambda_{1,t} - \beta E_t \lambda_{1,t+1} (1 - \gamma_{t+1} x_t^2) - \beta^2 E_t [((\beta b_t^\pi + \kappa)x_{t+1} + u_{t+1}) x_{t+1}] - \quad (18)$$

$$\beta^2 E_t [\lambda_{1,t+1} \gamma_{t+1} x_t x_{t+1}]$$

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

251 A key insight is that in the FOCs (17)-(18) all the terms that come from the
 252 manipulation of beliefs are weighted by the gain, and thus become irrelevant as
 253 $\gamma_t \rightarrow 0$, unless they grow unboundedly. In the Appendix we use this insight to
 254 rewrite the updating equations for beliefs as a stochastic recursive algorithm (SRA
 255 hereafter) in the standard form studied in Evans and Honkapohja [16]:

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

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

²⁹We 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.

³⁰The 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.

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

256 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 manip-
 257 ulation of beliefs are grouped in the second-order term ρ .³²

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

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

260 where the expectation is taken over the invariant distribution of the process $\widehat{Y}_t(\theta)$,
 261 which is the stochastic process for Y_t obtained by holding θ_{t-1} at the fixed value
 262 $\theta_{t-1} = \theta$.³³ Given the definition of \mathcal{H} provided in the Appendix, we get:

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

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

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

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

269 **Remark 1.** *The Jacobian (21) has negative eigenvalues for any value of the struc-*
 270 *tural parameters.*

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

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

³³It 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.

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

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

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

278 Proposition 2 implies that in the limit $\theta_t = [b_t^\pi, b_t^x]' \rightarrow 0$. This is the only possible
 279 E-stable equilibrium and it is locally stable. Equation (10) then shows that in the
 280 limit agents expect zero inflation and output gap. Substituting this together with
 281 $\gamma_t \rightarrow 0$ into the FOC (17) and the PC (2) implies that both output and inflation
 282 converge to the IT equilibrium (6)-(7).

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

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

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

³⁵From 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).

³⁶Strictly 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.

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

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

297 3. Policy Implications

298 Policy incentives behind our main result are best illustrated by the unfolding
299 dynamics. For presentational purposes, we will discuss simulations with constant
300 gain learning, because it allows us to focus on the policy trade-offs while abstracting
301 from the role of a changing gain parameter.³⁹ For our baseline simulations we set
302 $\gamma = 0.05$, which is a value consistent with estimates for the US economy⁴⁰, and
303 examine the role of the gain parameter at the end of Section 3.

304 3.1. Long- versus short-run policy trade-offs

305 Figure 1 illustrates our main result in welfare terms: as OP drives expectations
306 asymptotically to the IT equilibrium, expected welfare losses increase to those of IT.
307 For each time t , the figure plots the expected consumption equivalent (CE) measure of
308 welfare losses (percentage of steady-state consumption) for an economy starting from
309 period- t average beliefs; at time zero we start from PLT beliefs.⁴¹ For comparison
310 we plot the same CE measure for two Taylor-type rules, that Evans and Honkapohja
311 [18] and Evans and Honkapohja [17] have proven to drive beliefs respectively to PLT

³⁸In 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.

³⁹We 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.

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

⁴¹We 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.

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

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

317 The *long-run* benefits of PLT would be similar to the case with rational agents, i.e.
318 it anchors agents' inflation expectations once learning expectations have settled on
319 the equilibrium; "keeping" learning expectations in the PLT equilibrium is superior to
320 "keeping" them in the IT equilibrium. Similar results can be found also in different
321 setups, which all show that expectations are better anchored under PLT. Preston
322 [34] shows the robustness of long-term benefits of PLT to misinformation about
323 agents learning⁴³; in a framework featuring near-rational expectations, Woodford
324 [45] argues that benefits of engineering a history-dependent policy are present also
325 when expectations differ from RE with a nonspecified error structure.

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

⁴²The 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.

⁴³Preston [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.

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

337 Even though the CB takes advantage of its credibility during the transition, it has
338 *no incentive to build credibility* at any point in time. As the CB keeps engineering
339 surprise output contractions, expectations keep getting further away from PLT, and
340 agents believe less and less in a history-dependent policy (see Figure 2). OP is
341 however careful *not to lose credibility too fast*, in order to maintain its ability to
342 disinflate through lowering inflation expectations (i.e. keep $b^\pi > 0$, such that $\hat{E}\pi_{t+1} =$
343 $b^\pi x_t$ can be lowered by lowering x_t). Based on forecast errors, it would not be easy
344 for agents to conclude that the CB deviated from PLT (for more on this, see Section
345 5). First, they are small during the transition, similar in size to what would arise
346 in the PLT equilibrium (Figure 3).⁴⁵ Second, there is no systematic pattern in
347 forecast errors: agents sometimes overpredict, sometimes underpredict the outcome
348 (see Figure 4). Only when the economy converges close enough to IT do forecast
349 errors increase, as the CB loses its incentive to keep inflation expectations history
350 dependent. Where the CB really fools agents is in output expectations, but these
351 have a small impact on welfare losses.⁴⁶ As the economy converges on IT, forecast
352 errors become similar to those of a rational agent in IT. All these forecast errors are
353 however very small in magnitude.

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

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

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

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

361 5: for a bigger γ OP engineers less-aggressive output contractions in response to a
 362 positive cost-push shock.⁴⁷

363 The loss of credibility in the long run cannot be solved by delegation, in the
 364 spirit of Rogoff [36], by appointing a more patient central banker (higher β).⁴⁸ As
 365 long as future losses are discounted, $\beta < 1$, in the long run IT is the resulting
 366 equilibrium. We can observe in Figure 7b that all a more patient central banker
 367 achieves is keeping the economy close to the welfare-improving PLT equilibrium for
 368 a longer period, i.e. retaining “credibility” longer, because she is exploiting less the
 369 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

370 3.2. Short-run policy incentives

371 The short-run gains of OP come from the well-known time-inconsistency problem
 372 of PLT and the sluggishness of agents’ beliefs. The time inconsistency is standard:
 373 if given the chance, the CB has an incentive to renege its commitments and choose
 374 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

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

⁴⁸In 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.

⁴⁹A 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 ($\beta \rightarrow 1$) 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.

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

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

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

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

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

394 4. Implementation with a simple rule

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

398 This nonlinearity would also make its implementation impractical. Moreover, such a
 399 rule would require detailed knowledge of how agents learn. Such superior knowledge
 400 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 $\gamma = 0$.

$$i_t = \delta_\pi(b_{t-1}^\pi)E_t^*\pi_{t+1} + \delta_x E_t^*x_{t+1} + \delta_u(b_{t-1}^\pi)u_t, \quad (23)$$

where the coefficients are

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

401 This rule satisfies the Taylor principle, and guarantees both determinacy under
 402 RE and E-stability⁵⁰, hence yielding convergence under learning.

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

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

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

⁵¹Details available upon request.

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

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

418 5. Extensions

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

421 5.1. Generalizability

422 We conduct our analysis by assuming a specific learning algorithm. This algo-
423 rithm can be justified on several grounds⁵⁴ but it might seem arbitrary because it is
424 just one out of many. Yet, it is equally arbitrary to assume that there is no learning
425 component to private expectations, especially because this is at odds with recent
426 empirical findings.⁵⁵ Ultimately, how people form expectations is a yet unsettled is-
427 sue, which presents an important challenge for policymakers. Would they know the
428 exact expectation formation of agents, policy would be much easier. Then emerges
429 the question on what are the limits of our results, and whether they are robust to
430 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

⁵³We refer the reader to the Appendix for details.

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

⁵⁵See 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. $b_{t-1}^\pi x_{t-1}$ is positive, and π_t turns out to be even more positive, agents want to increase $b_t^\pi x_t$ to track π .

better forecasts can be found.⁵⁷

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

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

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

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

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

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

⁵⁷An 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.

⁵⁸Mathematically 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).

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

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

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

460 5.2. *Uncertainty about learning, and evolutionary dynamics*

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

468 Finally, in our main analysis we presumed little thinking on the agents side, while
 469 the policymaker is strategic.⁶⁰ This raises the question, whether agents would leave
 470 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).

⁶⁰See 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?”

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

476 6. Concluding remarks

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

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

⁶¹Note 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]).

496 rational and adaptive agents.

497 The CB incentives that arise in our framework have previously been ignored by
498 proponents of PLT under learning (see Evans and Honkapohja [18], Aoki and Nikolov
499 [2], Gaspar et al. [22]). Those authors showed that PLT is a learnable equilibrium:
500 if expectations are perturbed out of the PLT equilibrium, the CB can implement
501 a policy that makes agents learn the PLT equilibrium again. However, once CB
502 incentives are taken into account, PLT is no longer optimal if agents are learning.

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

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

526 **References**

- 527 [1] Adam, K., Marcet, A., 2011. Internal rationality, imperfect market knowledge
528 and asset prices. *Journal of Economic Theory* 146, 1224–1252.
- 529 [2] Aoki, K., Nikolov, K., 2006. Rule-Based Monetary Policy under Central Bank
530 Learning, in: *NBER International Seminar on Macroeconomics 2004*. National
531 Bureau of Economic Research, Inc. NBER Chapters, pp. 145–195.
- 532 [3] Barucci, E., Landi, L., 1997. Least mean squares learning in self-referential
533 linear stochastic models. *Economics Letters* 57, 313 – 317.
- 534 [4] Benveniste, A., Métivier, M., Priouret, P., 1990. *Adaptive Algorithms and
535 Stochastic Approximations*. Berlin: Springer-Verlag.
- 536 [5] Branch, W.A., 2004. The theory of rationally heterogeneous expectations: Ev-
537 idence from survey data on inflation expectations. *The Economic Journal* 114,
538 592–621.
- 539 [6] Branch, W.A., Evans, G.W., 2006. A simple recursive forecasting model. *Eco-
540 nomics Letters* 91, 158–166.
- 541 [7] Carvalho, C., Nechio, F., 2014. Do people understand monetary policy? *Journal
542 of Monetary Economics* 66, 108 – 123.
- 543 [8] Clarida, R., Gali, J., Gertler, M., 1999. The science of monetary policy: A new
544 keynesian perspective. *Journal of Economic Literature* 37, 1661–1707.
- 545 [9] Dal Bo, P., Frechette, G.R., 2011. The evolution of cooperation in infinitely
546 repeated games: Experimental evidence. *American Economic Review* 101, 411–
547 29.
- 548 [10] Del Negro, M., Eusepi, S., 2011. Fitting observed inflation expectations. *Journal
549 of Economic Dynamics and Control* 35, 2105–2131.
- 550 [11] Eggertsson, G.B., Woodford, M., 2003. The Zero Bound on Interest Rates and
551 Optimal Monetary Policy. *Brookings Papers on Economic Activity* 34, 139–235.

- 552 [12] Eusepi, S., Giannoni, M.P., Preston, B., 2018. Some implications of learning for
553 price stability. *European Economic Review* 106, 1 – 20.
- 554 [13] Eusepi, S., Preston, B., 2016. The science of monetary policy: an imperfect
555 knowledge perspective. Staff Reports. Federal Reserve Bank of New York.
- 556 [14] Evans, G.W., Honkapohja, S., 1998a. Convergence of learning algorithms with-
557 out a projection facility. *Journal of Mathematical Economics* , 59–86.
- 558 [15] Evans, G.W., Honkapohja, S., 1998b. Stochastic gradient learning in the cobweb
559 model. *Economics Letters* 61, 333 – 337.
- 560 [16] Evans, G.W., Honkapohja, S., 2001. *Learning and Expectations in Macroecon-*
561 *omics*. Princeton: Princeton University Press.
- 562 [17] Evans, G.W., Honkapohja, S., 2003. Expectations and the stability problem for
563 optimal monetary policies. *Review of Economic Studies* 70, 807–824.
- 564 [18] Evans, G.W., Honkapohja, S., 2006. Monetary policy, expectations and com-
565 mitment. *The Scandinavian Journal of Economics* 108, 15–38.
- 566 [19] Evans, G.W., Honkapohja, S., Williams, N., 2010. Generalized stochastic gra-
567 dient learning. *International Economic Review* 51, 237–262.
- 568 [20] Gaspar, V., Smets, F., Vestin, D., 2006a. Adaptive Learning, Persistence, and
569 Optimal Monetary Policy. *Journal of the European Economic Association* 4,
570 376–385.
- 571 [21] Gaspar, V., Smets, F., Vestin, D., 2006b. Optimal Monetary Policy under
572 Adaptive Learning. *Computing in Economics and Finance* 2006 183. Society for
573 Computational Economics.
- 574 [22] Gaspar, V., Smets, F., Vestin, D., 2007. Is time ripe for price level path stability?
575 Working Paper Series 0818. European Central Bank.

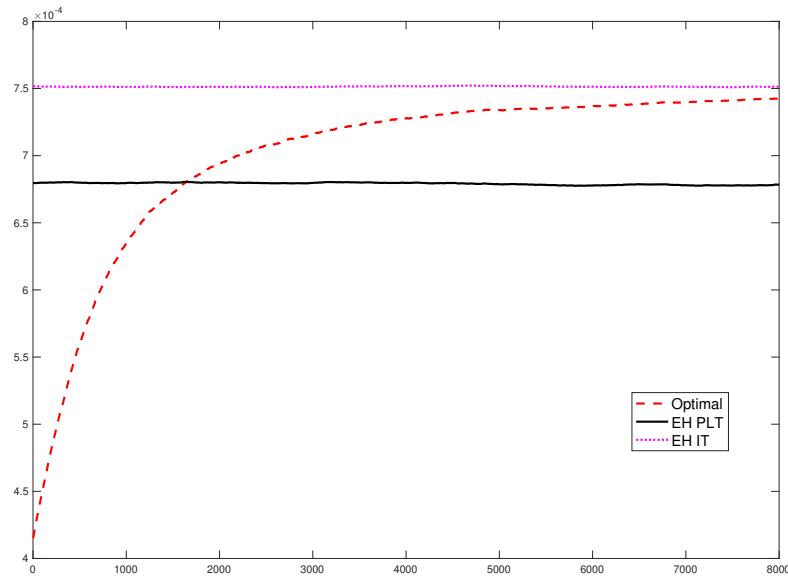
- 576 [23] Gorodnichenko, Y., Shapiro, M.D., 2007. Monetary policy when potential out-
577 put is uncertain: Understanding the growth gamble of the 1990s. *Journal of*
578 *Monetary Economics* 54, 1132–1162.
- 579 [24] Hatcher, M., Minford, P., 2016. Stabilisation policy, rational expectations and
580 price-level versus inflation targeting: a survey. *Journal of Economic Surveys* 30,
581 327–355.
- 582 [25] Honkapohja, S., Mitra, K., Evans, G.W., 2011. Notes on Agents’ Behavioral
583 Rules Under Adaptive Learning and Studies of Monetary Policy. SIRE Discus-
584 sion Papers 2011-04. Scottish Institute for Research in Economics (SIRE).
- 585 [26] Kurozumi, T., 2008. Optimal sustainable monetary policy. *Journal of Monetary*
586 *Economics* 55, 1277–1289.
- 587 [27] Lucas, R.E., 1987. *Models of Business Cycles*. Oxford: Basil Blackwell.
- 588 [28] Milani, F., 2006. A bayesian dsge model with infinite-horizon learning: Do
589 ”mechanical” sources of persistence become superfluous? *International Journal*
590 *of Central Banking* 2, 87–106.
- 591 [29] Milani, F., 2007. Expectations, learning and macroeconomic persistence. *Jour-*
592 *nal of Monetary Economics* 54, 2065–2082.
- 593 [30] Molnar, K., Ormeno, A., 2015. Using survey data of inflation expectations in
594 the estimation of learning and rational expectations models. *Journal of Money*
595 *Credit and Banking* 47, 673–699.
- 596 [31] Molnar, K., Santoro, S., 2014. Optimal monetary policy when agents are learn-
597 ing. *European Economic review* 66, 39–62.
- 598 [32] Preston, B., 2005. Learning about Monetary Policy Rules when Long-Horizon
599 Expectations Matter. *International Journal of Central Banking* 1.
- 600 [33] Preston, B., 2006. Adaptive learning, forecast-based instrument rules and mon-
601 etary policy. *Journal of Monetary Economics* 53, 507–535.

- 602 [34] Preston, B., 2008. Adaptive learning and the use of forecasts in monetary policy.
603 Journal of Economic Dynamics and Control 32, 3661 – 3681.
- 604 [35] Reis, R., 2013. Central bank design. Journal of Economic Perspectives 27,
605 17–44.
- 606 [36] Rogoff, K., 1985. The Optimal Degree of Commitment to an Intermediate
607 Monetary Target. The Quarterly Journal of Economics 100, 1169–89.
- 608 [37] Røisland, Ø., 2006. Inflation inertia and the optimal hybrid inflation/price-level
609 target. Journal of Money, Credit and Banking 38, 2247–2251.
- 610 [38] Rotemberg, J.J., Woodford, M., 1997. An optimization-based econometric
611 framework for the evaluation of monetary policy, in: Bernanke, B., Rotem-
612 berg, J.J. (Eds.), NBER Macroeconomics Annual 12. Cambridge (MA): MIT
613 Press.
- 614 [39] Sargent, T.J., 1999. The Conquest of American Inflation. Princeton: Princeton
615 University Press.
- 616 [40] Slobodyan, S., Wouters, R., 2012. Learning in an estimated medium-scale dsge
617 model. Journal of Economic Dynamics and Control 36, 26–46.
- 618 [41] Vestin, D., 2006. Price-level versus inflation targeting. Journal of Monetary
619 Economics 53, 1361–1376.
- 620 [42] Wolman, A.L., 2005. Real Implications of the Zero Bound on Nominal Interest
621 Rates. Journal of Money, Credit and Banking 37, 273–96.
- 622 [43] Woodford, M., 1999. Optimal monetary policy inertia. The Manchester School,
623 Supplement 67, 1–35.
- 624 [44] Woodford, M., 2003. Interest and Prices: Foundations of a Theory of Monetary
625 Policy. Princeton: Princeton University Press.

- 626 [45] Woodford, M., 2010. Robustly Optimal Monetary Policy with Near-Rational
627 Expectations. *American Economic Review* 100, 274–303.
- 628 [46] Yun, T., 1996. Nominal price rigidity, money supply endogeneity, and business
629 cycles. *Journal of Monetary Economics* 37, 345–370.

630 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

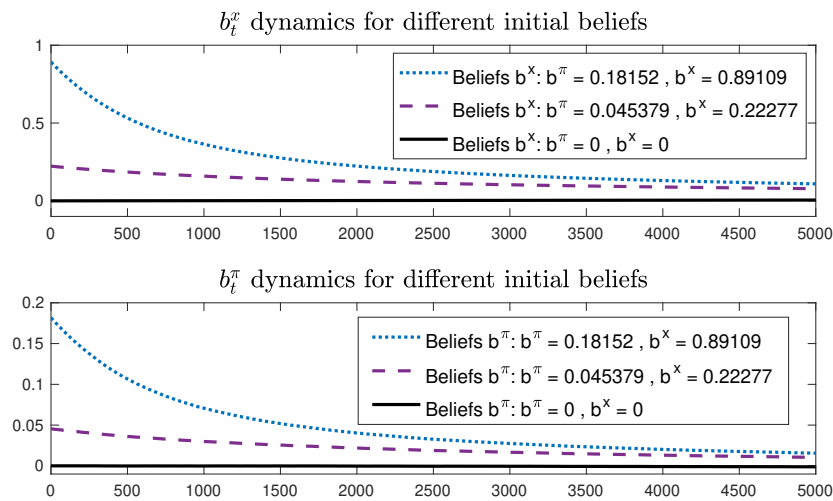


Figure 3: Squared forecast errors

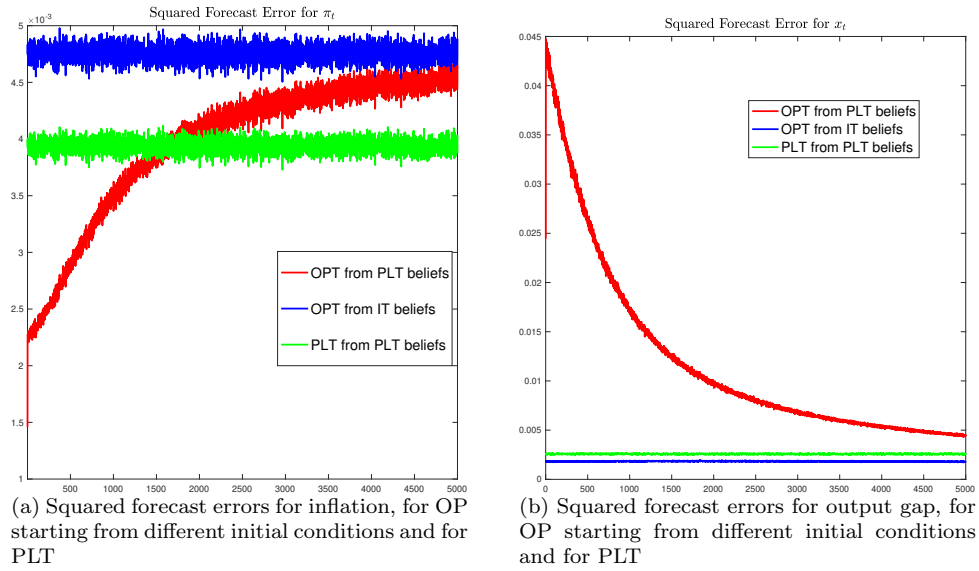


Figure 4: Forecast errors for output gap and inflation for one series, OP starting from PLT

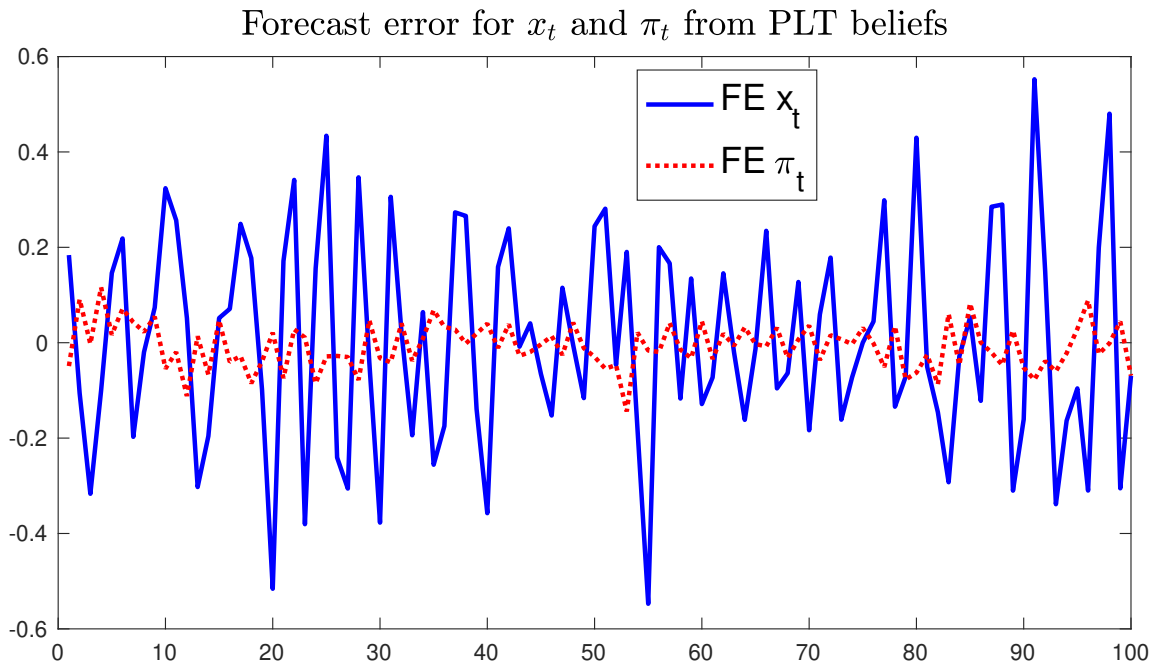


Figure 5: On-impact output gap responses with different private sector beliefs (to a one standard deviation cost-push shock)

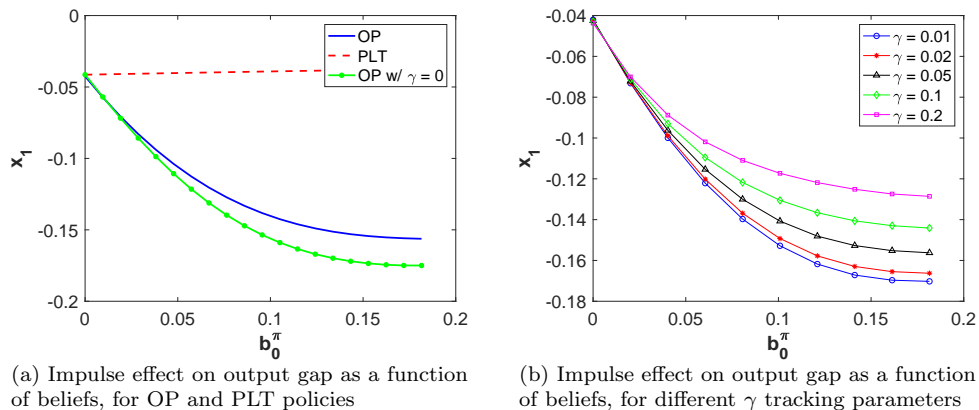


Figure 6: Impulse responses after a one standard deviation cost-push shock, under optimal policy under learning (OP) and price-level targeting policy (PLT), starting with initial beliefs corresponding to the rational expectations PLT equilibrium, with $\gamma = 0.05$.

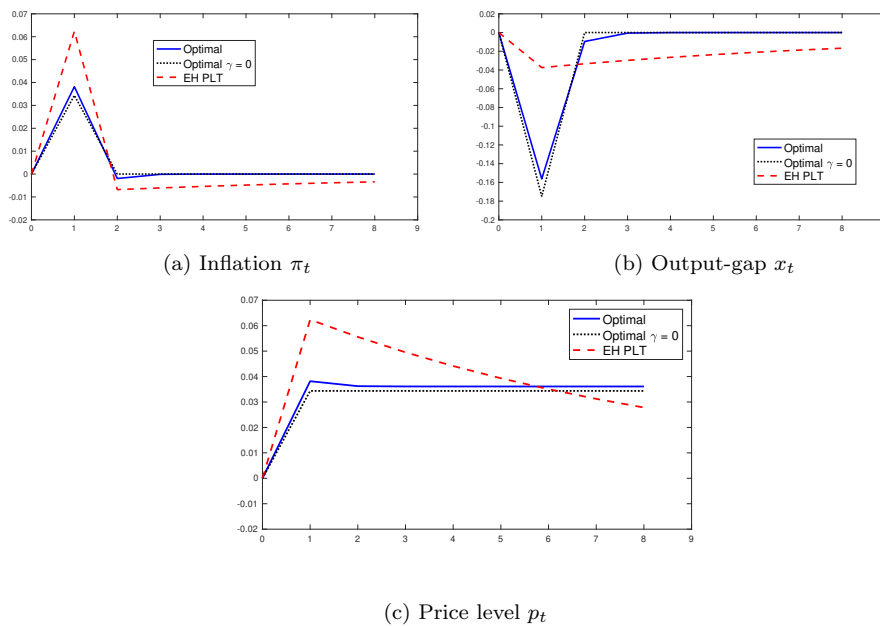
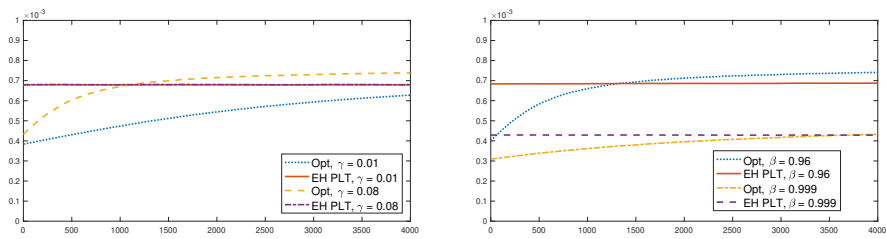


Figure 7: Consumption equivalents



(a) Slow and fast learning (low and high γ) (b) Patient and impatient CB (high and low β)