Discussion Papers in Economics

ON THE PERILS OF STABILIZING PRICES WHEN AGENTS ARE LEARNING

By

Antonio Mele
(University of Surrey)
Krisztina Molnár
(Norwegian School of Economics)
&
Sergio Santoro
(Bank of Italy)

DP 02/15

Department of Economics
University of Surrey
Guildford
Surrey GU2 7XH, UK
Telephone +44 (0)1483 689380
Facsimile +44 (0)1483 689548
Web www.econ.surrey.ac.uk
ISSN: 1749-5075

On the perils of stabilizing prices when agents are learning

Antonio Mele*, Krisztina Molnár† and Sergio Santoro‡

December 19, 2014

Abstract

We show that price level stabilization is not optimal in an economy where agents have incomplete knowledge about the policy implemented and try to learn it. A systematically more accommodative policy than what agents expect generates short term gains without triggering an abrupt loss of confidence, since agents update expectations sluggishly. In the long run agents learn the policy implemented, 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

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

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

‡Department of Economic Outlook and Monetary Policy Studies, Bank of Italy; Email: sergio.santoro@bancaditalia.it. A previous draft of this paper has been circulated under the title "The suboptimality of commitment equilbrium 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, Aarti Singh for useful comments. All the 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.

No monetary authority sets price level stabilization[1] as its official goal, despite economists' recommendation that this is the best way to conduct monetary policy. 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.[2] 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.[3]

This paper rationalizes why monetary authorities are so reluctant to implement price level stabilization. We examine the implications of this concern in a standard macroeconomic model, and we demonstrate that price level stabilization is not optimal if there is even the minimal chance that private sector misunderstands the policy regime.

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 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, Gali, and Gertler (1999) and Ambler (2009)) that price level stabilization

[1]Price level stabilization implies counteracting the effect of shocks on the price level.

[2]Swedish economists, like Gustav Cassel, David Davidson and Eli Heckscher held their firm support in public debates for price level targeting, 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.

[3]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."

is optimal.

We slightly depart from the assumption of rational expectations by postulating that agents do not know the exact mapping between shocks and aggregate variables induced by monetary policy.[4] We assume 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. The novelty of our setup is that a benevolent, fully rational monetary authority can "teach" agents the exact mapping by selecting an appropriate path for policy. In fact, the exact mapping is endogenous to policy choices. By choosing a particular policy response to shocks, the central bank affects agent's beliefs about the mapping. Those beliefs feed back into the evolution of aggregate variables, and thus into the mapping between shocks and aggregate variables. To find the optimal policy, we follow the methodology of Gaspar, Smets, and Vestin (2006) and Molnar and Santoro (2014), and assume that the central bank takes into account that its actions affect the data used in agents' estimations, and how those data affect their future beliefs.

Our main result is that price level stabilization is no longer optimal if agents are learning. We show that the policymaker wants to give up the benefits of stabilizing the price level in favour of short term gains.

The advantage of price level stabilisation arises from its history dependence: after a temporary shock that increases the price level, the policymaker should engineer 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

[4]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 (2011)). Moreover, an individual might be uncertain about other agents' knowledge about the exact mapping, which in turn would impact the evolution of aggregate variables (see Brock and Hommes (1997), Branch and McGough (2011), Molnar (2007)).

dependence, the policymaker can lower agents' expectations about future inflation by contracting current output.[5] Lower inflation expectations then decrease current inflation through the Phillips Curve.[6]

Under learning the central bank can attain short term gains because agents revise their beliefs very 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 amount of data to uncover 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. With such a policy, future output contractions are small or absent, and therefore they are not sufficient to bring the price level back to target. Hence, the price level rises permanently.

In the long run, monetary policy completely looses 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*. Under this policy, the central bank responds to shocks as long as they affect inflation. A temporary shock that increases the price level affects inflation on impact, but not in the future. Therefore the central bank counteracts the effect of the shock in the current period, but it does not spread it over future periods (see Gali (2003)). The long run policy recommendation is therefore in line with what many central banks set as their official goal.

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 central bank), which the previous literature did not consider. We therefore derive a novel convergence theorem that can accommodate the interaction between

[5]Evans and Honkapohja (2006) shows a policy that can convince learning agents that price level stabilization is in place. Note that, once agents have learned the mapping that would arise under rational expectations, the advantage of price level stabilization is similar under both learning and rational expectations.

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

updating rules for agents' beliefs and the choices of the rational central bank. This methodological contribution might be of separate interest to some readers, as our theorem and our line of proof could be applied in similar problems with a linear-quadratic setup.

There are several strands of existing literature on price level stabilization that are relevant to this paper. Many authors have shown its robustness: it anchors inflation expectations even if the central bank makes mistakes in forecasting output (Gorodnichenko and Shapiro (2007)) or faces model uncertainty (Aoki and Nikolov (2006)). By committing to a price level path, policy can alleviate the risks of hitting the zero lower bound (Eggertsson and Woodford (2003), Wolman (2005)). Contrary to these findings, our result is that price level stabilization is not the best policy if agents are not fully rational when facing a strategic central bank. Our results, however, do not call into question the long run advantages of price level stabilization. As we discussed before, in our model long run benefits arise purely from anchoring future inflation expectations. In a more general model, there are further advantages from the reduced long term variability of the price level for long-run nominal contracts and long run intertemporal decisions. For example Meh, Ros-Rull, and Terajima (2010) shows that price stabilization reduces long run redistributional effects from lenders to borrowers. Our result introduces an additional argument into this policy debate: the incentives of a rational policymaker change when there is even the smallest chance that agents could misunderstand policy choices.

This paper belongs to an extensive literature examining monetary policy when agents are learning. Bullard and Mitra (2002), Evans and Honkapohja (2003) and Bullard, Evans, and Honkapohja (2008), among others, show that policy rules that have good properties under rational expectations can have unintended and undesirable consequences if instead agents are learning. Our paper furthers this line of inquiry by considering how learning allows the central bank to do something better than price level stabilization, even if the latter remains a feasible and, in the long run, attractive strategy.

Our work is also related to a wider literature that proposes learning as a useful

way to evaluate and modify the traditional equilibrium concepts.[7] Learning mechanisms proposed by the literature range from simple rules of thumb, to more sophisticated rules like Bayesian learning (see for example Beggs (2005) and Borgers and Sarin (1997)) and adaptive learning (i.e. learning with econometric methods), like the one we use in this paper. Adaptive learning is especially useful when agents learn about a self referential variable (i.e. one which depends on the agents' actions), mostly because in this case Bayesian learning rules are intractable.[8] In self referential models, least squares learning has long been used for refining rational expectation equilibria. Several authors in particular have used least squares learning for equilibrium selection, and for asking how policy can guarantee a learnable equilibrium (see, among others Eusepi and Preston (2010), Marcet and Sargent (1989a), Marimon and Sunder (1993), Adam (2003), Bullard and Mitra (2002), and Evans and Honkapohja (2001) for an extensive survey). This paper refines the existing concept of learnability, by taking into account strategic interaction among players with different expectations formation mechanisms. Our model features two rational expectations equilibria which are both learnable; yet, the incentives of an optimizing rational agent eliminate one of the learnable equilibria.

Our analysis highlights an important message about adaptive learning: even if agents learn rational expectations equilibria and their forecasts cannot be distinguished from a rational agent, they do not form strategies like a rational player. Therefore, a rational agent facing learners will not behave in the same way as when facing rational agents. When a rational policymaker faces rational agents, a deviation from the price stabilizing policy would be immediately realized by agents, who in turn would change their beliefs abruptly and assume the central bank is following an alternative policy. This off-equilibrium threat of rational agents can keep the central bank from deviating from the price stabilizing policy (see Kurozumi (2008)). 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 central bank. This lack of off-equilibrium strategies

[7] A learning model in the broad sense is "any model that specifies the learning rules used by individual players, and examines their interaction"(Fudenberg and Levine (1998) p3).

[8] A rational Bayesian learner would understand how its 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.

provides strong incentives for the rational policymaker to deviate from the price stabilization policy.

1 The Model

We consider the baseline version of the New Keynesian model, and as it is standard, we log-linearize the equilibrium equations and take a second-order Taylor approximation of the agent's utility function. The economy is therefore characterized by two structural equations.[9] The first one is an IS equation:

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

where x_t, r_t and π_t denote the time t output gap (i.e. the difference between actual and natural output), the short-term nominal interest rate and inflation, respectively; σ is a parameter of the household's utility function, representing risk aversion. Note that the operator E_t^* represents agents' conditional expectations, which are not necessarily rational. The above equation is derived by log-linearizing the household's Euler equation and imposing the equilibrium condition that consumption equals output.

The second equation is the so-called New Keynesian Phillips Curve (NKPC):

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

where β denotes the subjective discount rate, κ is a function of structural parameters, and $u_t \sim N(0, \sigma_u^2)$ is a white noise cost-push shock[10]; this relation is obtained from optimal pricing decisions of monopolistically competitive firms whose prices are staggered à la Calvo (1983).[11]

[9] For details of the derivation of the structural equations of the New Keynesian model see, among others, Yun (1996), Clarida, Gali, and Gertler (1999) and Woodford (2003).

[10] Note that the cost-push shock is usually assumed to be an AR(1) process, however we instead assume it to be *iid* to make the problem more tractable. This assumption is also supported by Milani (2006), who shows that learning can endogenously generate persistence in inflation data, and assuming a strongly autocorrelated cost-push shock becomes redundant.

[11] In other words, the probability that a firm in period t can reset the price is constant over time and across firms.

The central bank (CB in short) is benevolent and therefore acts as the social planner. It then maximizes the agents' utility function subject to the structural equations described above. By deriving a second-order approximation for the utility function, we can express the objective of the central bank as a loss function in the following form:

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

where α is the relative weight put by the CB on the objective of output gap stabilization.[12]

1.1 Price level targeting vs inflation targeting under RE

Assume that the private sector has rational expectations (RE in short), and that the CB can credibly commit to a future course of action. The policy problem is to minimize the social welfare loss (3), subject to the structural equations (1) and (2), where E_t^* is replaced by E_t:

$$\begin{aligned}&\min_{\{\pi_t,x_t,r_t\}_{t=0}^{\infty}} E_0\sum_{t=0}^{\infty}\beta^t\left(\pi_t^2+\alpha x_t^2\right) \\ &\text{s.t.(1), (2)}\end{aligned} \tag{4}$$

As shown, among others, in Clarida, Gali, and Gertler (1999), the optimality conditions of this problem are:

$$\pi_0 = -\frac{\alpha}{\kappa}x_0 \tag{5}$$

$$\pi_t = -\frac{\alpha}{\kappa}x_t+\frac{\alpha}{\kappa}x_{t-1}, \quad t\geq 1 \tag{6}$$

[12] Rotemberg and Woodford (1997) show how (3) can be obtained as a quadratic approximation to the expected household's utility function. The parameter α is a function of structural parameters.

Hence, the optimality condition at time 0 is different from that holding at $t \geq 1$. The term in x_{t-1} that appears when $t \geq 1$ represents the past promises that the CB committed to realize at time t; hence, is absent for $t = 0$, when there are no promises to be kept. A policy characterized by the equations (5)-(6) is prone to time inconsistency: if the policymaker could reoptimize at a date $T > 0$, the optimality condition at T would be different from that implied by (6). We follow Woodford (2003)'s "timeless perspective" and use (6) as the only relevant optimality condition.

Combining (6) with the NKPC (2), Clarida, Gali, and Gertler (1999) shows that output gap and inflation evolve according to the following law of motion:

$$x_t = b^x x_{t-1} + c^x u_t \tag{7}$$

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

where the coefficients are given by:

$$b^x = \frac{\kappa^2 + \alpha(1+\beta) - \sqrt{(\kappa^2 + \alpha(1+\beta))^2 - 4\alpha^2\beta}}{2\alpha\beta} \tag{9}$$

$$b^\pi = \frac{\alpha}{\kappa}(1 - b^x) \tag{10}$$

$$c^x = -\frac{\kappa b^x}{\alpha} \tag{11}$$

$$c^\pi = -\frac{\alpha}{\kappa} c^x \tag{12}$$

Clarida, Gali, and Gertler (1999) show that the policy implied by (7)-(8) is equivalent to price level targeting (PLT in short): the central bank responds to changes in the price level, and tries to keep prices close to a predetermined value.

Now assume the central bank cannot commit to future policy, and therefore it acts discretionarily when a shock hits the economy. In this case, the monetary authority solves the problem 4 by taking future expected policy as given. Clarida, Gali, and Gertler (1999) shows that the optimal allocation obeys the following equation

$$\pi_t = -\frac{\alpha}{\kappa} x_t \tag{13}$$

Using the NKPC (2), it is easy to show that output gap and inflation are charac-

terized by

$$x_t = -\frac{\kappa}{\alpha + \kappa^2} u_t \tag{14}$$

$$\pi_t = \frac{\alpha}{\alpha + \kappa^2} u_t \tag{15}$$

We call this inflation targeting (IT in short), since as shown in Clarida, Gali, and Gertler (1999) the central bank responds to changes in inflation, 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 (1999): the current allocations depend on past levels of output gap. At the contrary, the IT policy only depends on current shocks.

1.2 Learning specification

In the rest of the paper, we dispose of the assumption that the private sector has RE. Following Molnar and Santoro (2014), we posit that the central bank is fully rational. However, we assume that agents are adaptive learners. This assumption postulates that agents know the structure of the economy, and they are able to calculate the rational expectations equilibrium. However, they are uncertain about some parameters' values. Hence, they estimate equilibrium conditions by observing past and current allocations.[13]

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 central bank followed the PLT policy under RE, i.e. (7)-(8). Hence, the PLM is:

$$\pi_t = b^{\pi} x_{t-1} + c^{\pi} u_t \tag{16}$$

$$x_t = b^{x} x_{t-1} + c^{x} u_t, \tag{17}$$

Under learning, agents estimate the coefficients in equations (16)-(17), and use

[13]The modern literature on adaptive learning was initiated by Marcet and Sargent (1989b), who were the first to apply stochastic approximation techniques to study the convergence of learning algorithms. For an extensive monograph on this paradigm, see Evans and Honkapohja (2001).

their estimates of b^{π}_{t-1} and b^{x}_{t-1} to make forecasts:

$$E^{*}_{t}\pi_{t+1} = b^{\pi}_{t-1}x_{t} \tag{18}$$

$$E^{*}_{t}x_{t+1} = b^{x}_{t-1}x_{t} \tag{19}$$

Notice that equations (16)-(17) are consistent with both PLT and IT policies. Hence, this specification allows agents to potentially learn both those policies. Intuitively, if the central bank consistently implements a PLT policy, agents would learn this policy. On the other hand, if the central bank consistently implements the IT policy, agents' beliefs about equations (16)-(17) will eventually be consistent with an IT policy. In other words, the model that agents estimate is consistent with both policies, and hence the central bank can potentially make them learn one or the other.

In the above equations we are assuming that x_t is part of the time t information set of the agents. This introduces a simultaneity problem between $E^{*}_{t}y_{t+1}$ and y_t that complicates the analysis of asymptotic convergence of the beliefs. In the learning literature this simultaneity problem is often solved by adopting a different timing convention, such that realized values of the endogenous variables y are included in the time t information set only up to time $t-1$. However, this alternative information assumption would increase the dimension of the state space: the forecasts of π_{t+1} and x_{t+1} would become:

$$E^{*}_{t}\pi_{t+1} = b^{\pi}_{t-1}\left(b^{x}_{t-1}x_{t-1} + c^{x}_{t-1}u_{t}\right) \tag{20}$$

$$E^{*}_{t}x_{t+1} = b^{x}_{t-1}\left(b^{x}_{t-1}x_{t-1} + c^{x}_{t-1}u_{t}\right). \tag{21}$$

Since expectations depend also on the estimated values of the coefficients c^{π} and c^{x}, an optimizing CB should take those (and their recursive estimation algorithm) into account. The central bank problem would then have two more state variables, with significant additional complications in the numerical exercise. To avoid this complications, we assume that agents' estimates are obtained with stochastic gradient learning. This assumption substantially implies that we can abstract from the evolution of the estimated second moments of the regressors, and hence forget about c^{π} and c^{x}. The recursive updating formula for the remaining estimated

coefficients is then

$$b_t^\pi = b_{t-1}^\pi + \gamma_t x_{t-1}\left(\pi_t - x_{t-1} b_{t-1}^\pi\right) \tag{22}$$

$$b_t^x = b_{t-1}^x + \gamma_t x_{t-1}\left(x_t - x_{t-1} b_{t-1}^x\right), \tag{23}$$

where γ_t is the so called gain parameter. When deriving our analytical results, we use $\gamma_t = \frac{1}{t}$ (in the literature this is called decreasing gain learning). For the numerical exercises, we use $\gamma_t = \gamma$ for some small number γ (this is defined as constant gain learning). The latter is done for presentational purposes only, and numerical results with decreasing gain are available upon request.

2 Optimal monetary policy

In this section, we derive the optimal monetary policy and prove the main convergence result. To ease analytical tractability, we assume agents follow decreasing gain learning, so that their estimates can eventually settle down to a limit point. Since the dynamic problem is non-standard, we first show that it has a recursive formulation where the state variables are the output gap, the parameters of the PLM, and the gain parameter. We then show that under the optimal policy, the IT equilibrium is stable under learning.

2.1 Recursivity

We start stating the control problem of the central bank in the case of decreasing gain. We write it as a maximization (instead of a minimization) problem, in order to refer more directly to the dynamic programming results.

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

s.t.

$$\begin{aligned}
x_t &= \frac{-\sigma^{-1}r_t}{1-b_{t-1}^x-\sigma^{-1}b_{t-1}^\pi}\\
\pi_t &= (\beta b_{t-1}^\pi+\kappa)x_t+u_t\\
b_t^\pi &= b_{t-1}^\pi+\gamma_t x_{t-1}\left(\pi_t-x_{t-1}b_{t-1}^\pi\right)\\
b_t^x &= b_{t-1}^x+\gamma_t x_{t-1}\left(x_t-x_{t-1}b_{t-1}^x\right),\\
&x_{-1},b_{-1}^\pi,\,b_{-1}^x,\gamma_0 \text{ given}
\end{aligned}$$

Since the IS curve is never a binding constraint (the central bank can always choose an interest rate that satisfy it, given the allocations and the beliefs), we can dispense from it. Using the NKPC to substitute out π the problem can be written in a simpler form:

$$\sup_{\{x_t,b_t^\pi,b_t^x\}_{t=0}^\infty} E_0(1-\beta)\sum_{t=0}^{\infty}\beta^t\left\{-\frac{1}{2}\left[\left((\beta b_{t-1}^\pi+\kappa)x_t+u_t\right)^2+\alpha x_t^2\right]\right\} \tag{24}$$

s.t.

$$b_t^\pi = b_{t-1}^\pi+\gamma_t x_{t-1}\left((\beta b_{t-1}^\pi+\kappa)x_t+u_t-x_{t-1}b_{t-1}^\pi\right) \tag{25}$$

$$b_t^x = b_{t-1}^x+\gamma_t x_{t-1}\left(x_t-x_{t-1}b_{t-1}^x\right), \tag{26}$$

$$x_{-1},b_{-1}^\pi,\,b_{-1}^x,\gamma_0 \text{ given} \tag{27}$$

There are five state variables. Three are endogenous $(x_{t-1},b_{t-1}^\pi,b_{t-1}^x)$, and take values in $\mathbb{R}^3$. One is exogenous and stochastic (u_t), defined over some underlying probability space, and takes values in a measurable space $(Z,\mathfrak{Z})$. Finally, there is one exogenous and deterministic state (γ_t) that takes values in a countable set $G\subset[0,1]$ and evolves following the recursion $\frac{1}{\gamma_t}=\frac{1}{\gamma_{t-1}}+1$. We denote the state space $S\equiv\mathbb{R}^3\times Z\times G$. The actions decided by the central bank are three (x_t,b_t^π,b_t^x); we denote this vector as a and the action space is $\mathbb{R}^3$. The feasibility

correspondence $\Gamma : S \to \mathbb{R}^3$ is defined as follows:

$$\text{for any } s \in S,\ \Gamma(s) = \left\{ a \in \mathbb{R}^3 : \text{equations (25) and (26) hold} \right\}$$

This optimization problem has some non-standard features. First of all, the graph of the feasibility correspondence is not convex, which implies that usual tools of concave programming cannot be used. Moreover, Γ is not compact-valued. Finally, the quadratic return function is unbounded below. For these reasons, in the statement of the problem we used the sup operator instead of the max, since the existence of a maximizing plan cannot be taken for granted.

We aim at proving that there exists an optimal time-invariant policy function that maximizes the objective function in (24). To do so, the strategy we adopt is the following: we write down a new maximization problem augmented by some arbitrary constraints that guarantee that the feasibility correspondence is compact-valued, and show that in this case there exists a time-invariant optimal policy function; then, we argue that these arbitrary constraints can be chosen so that they don't bind in an optimum, and that no optimum of the original problem can lie outside these constraints. Hence, we conclude that the standard FOCs can be used to characterize the optima of the original problem.

Note that 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.

We now write the new optimization problem:

$$\sup_{\{x_t, b_t^\pi, b_t^x\}_{t=0}^{\infty}} E_0(1-\beta) \sum_{t=0}^{\infty} \beta^t \left\{ -\frac{1}{2} \left[\left((\beta b_{t-1}^\pi + \kappa) x_t + u_t \right)^2 + \alpha x_t^2 \right] \right\} \tag{28}$$

s.t.

$$b_t^\pi = b_{t-1}^\pi + \gamma_t x_{t-1} \left((\beta b_{t-1}^\pi + \kappa) x_t + u_t - x_{t-1} b_{t-1}^\pi \right) \tag{29}$$

$$b_t^x = b_{t-1}^x + \gamma_t x_{t-1} \left(x_t - x_{t-1} b_{t-1}^x \right), \tag{30}$$

$$\overline{x}(s_t) \geq x_t \geq -\overline{x}(s_t), \tag{31}$$

$$x_{-1}, b_{-1}^\pi, b_{-1}^x, \gamma_0 \text{ given} \tag{32}$$

where we used the arbitrary continuous function of the states $\overline{x}(s_t)$. Let's now fix some notation. The vector of the state variables at time t is $s_t \equiv [x_{t-1}, b^{\pi}_{t-1}, b^{x}_{t-1}, u_t, \gamma_t]'$, while the vector of choice variables at t is $a_t \equiv [x_t, b^{\pi}_t, b^{x}_t]'$. We denote with a superscript i the i-th element of a vector. Hence, the evolution of the state variables can be summarized as follows:

$$\begin{aligned} s^1_{t+1} &= a^1_t \\ s^2_{t+1} &= a^2_t \\ s^3_{t+1} &= a^3_t \\ s^4_{t+1} &= \xi \\ s^5_{t+1} &= \frac{s^5_t}{1+s^5_t} \end{aligned}$$

where ξ is the realization of a random variable with the same distribution as u. We can represent the above relations in a more compact way:

$$s_{t+1} = \Psi(s_t, a_t, \xi) \tag{33}$$

Note that the operator Ψ is trivially continuous.

The transition probability from the graph of the feasibility correspondence to a Borel set $D \subset S$ is defined as:

$$Q(D|s,a) = \int_Z \mathbf{1}_D(\Psi(s,a,\xi))\, dP(\xi) \tag{34}$$

where $\mathbf{1}_D$ is the indicator function relative to set D, and P is the probability distribution of ξ.

We can now state and prove this simple Lemma.

Lemma 1. *The following results hold:*

(i) *The feasibility correspondence:*

$$\textit{for any } s \in S,\ \Gamma^c(s) = \left\{ a \in \mathbb{R}^3 : \textit{equations } (29),\ (30) \textit{ and } (31) \textit{ hold} \right\}$$

is compact-valued.

(ii) *The feasibility correspondence:*

$$\textit{for any } s \in S,\ \Gamma^c(s) = \left\{ a \in \mathbb{R}^3 : \textit{equations } (29),\ (30) \textit{ and } (31) \textit{ hold} \right\}$$

is upper hemi-continuous.

(iii) *For any bounded continuous function $v : S \to \mathbb{R}$, the function:*

$$F(s,a) = \int_S v(y)\, Q(dy|s,a)$$

is continuous.

Proof. **(i)** For any value of $s \in S$, equation (29) is a linear function of b_t^π and x_t, and analogously equation (30) is a linear function of b_t^x and x_t. Moreover, define:

$$\begin{aligned}\overline{b}^\pi(s_t) = \max\big\{ & b_{t-1}^\pi + \gamma_t x_{t-1}\left((\beta b_{t-1}^\pi + \kappa)\overline{x}(s_t) + u_t - x_{t-1} b_{t-1}^\pi\right), \\ & b_{t-1}^\pi + \gamma_t x_{t-1}\left((\beta b_{t-1}^\pi + \kappa)(-\overline{x}(s_t)) + u_t - x_{t-1} b_{t-1}^\pi\right)\big\}\end{aligned}$$

and:

$$\begin{aligned}\underline{b}^\pi(s_t) = \min\big\{ & b_{t-1}^\pi + \gamma_t x_{t-1}\left((\beta b_{t-1}^\pi + \kappa)\overline{x}(s_t) + u_t - x_{t-1} b_{t-1}^\pi\right), \\ & b_{t-1}^\pi + \gamma_t x_{t-1}\left((\beta b_{t-1}^\pi + \kappa)(-\overline{x}(s_t)) + u_t - x_{t-1} b_{t-1}^\pi\right)\big\}\end{aligned}$$

and analogously for $\overline{b}^x(s_t)$ and $\underline{b}^x(s_t)$. Hence, it is clear that:

$$\Gamma^c(s) \subset [-\overline{x}(s), \overline{x}(s)] \times [\underline{b}^\pi(s), \overline{b}^\pi(s)] \times [\underline{b}^x(s), \overline{b}^x(s)] \tag{35}$$

Moreover, by linearity (conditional on s) of the equations (29) and (30), we can argue that $\Gamma^c(s)$ is closed; since it is a closed subset of a compact set, we conclude that it is compact. Since s is arbitrary, Γ^c is compact-valued.

(ii) Let's consider an arbitrary sequence $\{s_n\}$ with $s_n \in S$ for any n, converging to a point $\widehat{s}$, and an arbitrary sequence $\{x_n\}$ with $x_n \in [-\overline{x}(s_n), \overline{x}(s_n)]$.

Then by continuity of $\overline{x}(\cdot)$ it is easy to show that there exists a convergent subsequence $\{x_{n_k}\}$ whose limit is in $[-\overline{x}(\widehat{s}), \overline{x}(\widehat{s})]$; moreover, the functional form of (29) and (30) (they are formed by sums and products of elements of $\{s_n\}$ and $\{x_n\}$) implies that if the subsequences $\left\{b^{\pi}_{n_k}\right\}$ and $\left\{b^{x}_{n_k}\right\}$ satisfy equations (29) and (30) for any n_k, then they converge and the limit satisfies (29) and (30) evaluated in the limits of $\{s_{n_k}\}$ and $\{x_{n_k}\}$. Since the sequences $\{s_n\}$ and $\{x_n\}$ are arbitrary, upper hemi-continuity of Γ^c is proved.

(iii) Consider an arbitrary sequence $\{s_n, a_n\}$ with $(s_n, a_n) \in S \times \mathbb{R}^3$ for any n, converging to a limit $(\overline{s}, \overline{a}) \in S \times \mathbb{R}^3$. We can use the Bounded Convergence Theorem (remember that the function v is bounded by assumption), continuity of v and Ψ and equation (34) to claim that:

$$\begin{aligned}
\lim_{n\to\infty} F(s_n, a_n) &= \lim_{n\to\infty} \int_S v(y)\, Q(dy|s_n, a_n) = \lim_{n\to\infty} \int_Z v(\Psi(s_n, a_n, \xi))\, dP(\xi) \\
&= \int_Z \lim_{n\to\infty} v(\Psi(s_n, a_n, \xi))\, dP(\xi) = \int_Z v(\Psi(s, a, \xi))\, dP(\xi) \\
&= F(s, a)
\end{aligned}$$

Since the sequence $\{s_n, a_n\}$ is arbitrary, continuity of F is proved.

□

We are now ready to prove the following Proposition.

Proposition 1. *There exists a time-invariant policy function for the CB that solves the optimization problem 28.*

Proof. This result follows from Theorem 1 of Jaskiewicz and Nowak (2011).[14] The assumptions of their Theorem are satisfied in our setup; most of them are proved in our Lemma 1, while the existence of a one-sided majorant function that satisfies their conditions (M1) and (M2) (see the Appendix for their exact formulation) is trivial in our model: since the quadratic return function of the CB is non-positive, a constant function $\omega(s) = 1$ for any $s \in S$ has the required properties.

Finally, note that their Theorem is derived in the case of a maxmin problem of a controller in a two-players game; assuming that the second player can play only

[14] We report the statement of the Theorem and its assumptions in the Appendix.

one strategy allows us to apply their results to our model. □

Next, we prove that any optimal time-invariant policy function for the problem 28 is such that the constraint (31) never binds in the optimum, if an appropriate continuous function $\overline{x}(s)$ is chosen. We define $V^c(s)$ as the value function associated with the solution of the problem 28 for a given initial vector of states $s \in S$.[15] In the following simple Lemma we characterize bounds of this value function.

Lemma 2. *Assume that the shock u has finite variance σ_u^2. The following results hold:*

(i) *For any $s \in S$ and any choice of $\overline{x}(s)$:*

$$V^c(s) \leq 0$$

(ii) *For any $s \in S$ and any choice of $\overline{x}(s)$:*

$$V^c(s) \geq -\frac{1}{2}\left[(1-\beta)u^2 + \beta\sigma_u^2\right]$$

where u is the fourth component of the vector s of initial states.

Proof. **(i)** This follows trivially from the fact that the one-period return function of the CB is non-positive.

(ii) For any choice of $\overline{x}(s)$, the allocation $x_t = 0$ for any $t \geq 0$ and any history of states is always feasible; with this allocation the welfare of the CB is given by:

$$E_0(1-\beta)\sum_{t=0}^{\infty}\beta^t\left\{-\frac{1}{2}\left[\left((\beta b_{t-1}^{\pi}+\kappa)x_t+u_t\right)^2+\alpha x_t^2\right]\right\} =$$
$$E_0(1-\beta)\sum_{t=0}^{\infty}\beta^t\left\{-\frac{1}{2}(u_t)^2\right\} = -\frac{1}{2}\left[(1-\beta)u_0^2+\beta\sigma_u^2\right]$$

Hence, the optimal allocation cannot deliver a welfare smaller than the one associated with this feasible allocation.

[15] Note that this value function depends also on the choice of $\overline{x}_s$, even if we do not make this dependence explicit.

□

We can now state and prove the following Proposition.

Proposition 2. *Let* $\overline{x}(s) = \epsilon\sqrt{\frac{(1-\beta)u^2+\beta\sigma_u^2}{\alpha(1-\beta)}}$*, for some* $\epsilon > 1$*; then any optimal time-invariant policy function for the problem 28 is such that the constraint* (31) *never binds.*

Proof. Theorem 1 of Jaskiewicz and Nowak (2011) shows that there exists a recursive formulation of our maximization problem, which is the following:

$$V^c(s) = -(1-\beta)\frac{1}{2}\left[(\beta b^{\pi}+\kappa)x^*(s)+u\right)^2 + \alpha x^{*2}(s)\right] + \beta\int_S V^c(s)\,Q\left(dy|s,a^*(s)\right) \tag{36}$$

for any $s \in S$, where the starred variables denote actions taken under any optimal policy function. Using Lemma 2 (i) and the fact that $-(1-\beta)\frac{1}{2}\left(\beta b^{\pi}+\kappa)x^*(s)+u\right)^2$ is non-positive, we have that:

$$V^c(s) \leq -(1-\beta)\frac{1}{2}\alpha x^{*2}(s)$$

Now, for the sake of contradiction, let's assume that for some $s \in S$ we have that $x^*(s) = \overline{x}(s)$.[16] This means that:

$$-x^{*2}(s) < -\frac{(1-\beta)u^2+\beta\sigma_u^2}{\alpha(1-\beta)}$$

which implies:

$$V^c(s) \leq -(1-\beta)\frac{1}{2}\alpha x^{*2}(s) < -\frac{1}{2}\left[(1-\beta)u^2+\beta\sigma_u^2\right] \tag{37}$$

which contradicts Lemma 2 (ii). □

2.2 Convergence

So far we proved that there exists an optimal time-invariant solution to the problem 28 and that it is interior; hence, any such solution can be characterized as

[16] We can proceed analogously for the case $x^*(s) = -\overline{x}(s)$.

the solution of the standard FOCs, without having to worry about the Lagrange multipliers on the constraints (31). The first order conditions of problem 28 are:

$$\begin{aligned} 0 =& -\alpha x_t - \left[(\beta b^{\pi}_{t-1} + \kappa)x_t + u_t\right](\beta b^{\pi}_{t-1} + \kappa) - \lambda_{1,t}\gamma_t x_{t-1}(\beta b^{\pi}_{t-1} + \kappa) - \\ & - E_t[\lambda_{1,t+1}\beta\gamma_{t+1}((\beta b^{\pi}_t + \kappa)x_{t+1} + u_{t+1} - b^{\pi}_t 2x_t)] - \lambda_{2,t}\gamma_t x_{t-1} \\ & - E_t[\lambda_{2,t+1}\beta\gamma_{t+1}(x_{t+1} - b^{x}_t 2x_t)] \end{aligned} \tag{38}$$

$$\begin{aligned} 0 =& \lambda_{1,t} - \beta E_t\lambda_{1,t+1}(1 - \gamma_{t+1}x_t^2) - \beta^2 E_t[((\beta b^{\pi}_t + \kappa)x_{t+1} + u_{t+1})\, x_{t+1}] - \\ & \beta^2 E_t\left[\lambda_{1,t+1}\gamma_{t+1}x_t x_{t+1}\right] \end{aligned} \tag{39}$$

$$0 = \lambda_{2,t} - \beta E_t\lambda_{2,t+1}(1 - \gamma_{t+1}x_t^2), \tag{40}$$

where $\lambda_{1,t}$ and $\lambda_{2,t}$ are the Lagrange multipliers of (29) and (30), respectively. These first order conditions together with the law of motion for the learning coefficients constitute the necessary conditions for the optimal evolution of $\{x_t, b^{\pi}_t, b^{x}_t\}$.[17] From equation (38) it is easy to show that the only stationary solution for $\lambda_{2,t}$ is $\lambda_{2,t} = 0$ for any t; hence the FOCs can be rewritten as:

$$\begin{aligned} 0 =& -\alpha x_t - \left[(\beta b^{\pi}_{t-1} + \kappa)x_t + u_t\right](\beta b^{\pi}_{t-1} + \kappa) - \lambda_{1,t}\gamma_t x_{t-1}(\beta b^{\pi}_{t-1} + \kappa) - \\ & - E_t[\lambda_{1,t+1}\beta\gamma_{t+1}((\beta b^{\pi}_t + \kappa)x_{t+1} + u_{t+1} - b^{\pi}_t 2x_t)] \end{aligned} \tag{41}$$

$$\begin{aligned} 0 =& \lambda_{1,t} - \beta E_t\lambda_{1,t+1}(1 - \gamma_{t+1}x_t^2) - \beta^2 E_t[((\beta b^{\pi}_t + \kappa)x_{t+1} + u_{t+1})\, x_{t+1}] - \\ & \beta^2 E_t\left[\lambda_{1,t+1}\gamma_{t+1}x_t x_{t+1}\right] \end{aligned} \tag{42}$$

Remembering that by Proposition 1 we can concentrate on time-invariant laws of motion for the optimal x, we can rewrite equation (41) as:

$$x_t = \Phi_1\left(b^{\pi}_{t-1}\right)u_t + \Phi_2\left(s_t\right) \tag{43}$$

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

where the vector s_t is the vector of state variables defined above, and:

$$\Phi_1\left(b_{t-1}^{\pi}\right) \equiv -\frac{\beta b_{t-1}^{\pi}+\kappa}{\alpha+\left(\beta b_{t-1}^{\pi}+\kappa\right)^2} \tag{44}$$

$$\begin{aligned}\Phi_2\left(s_t\right) \equiv & -\frac{1}{\alpha+\left(\beta b_{t-1}^{\pi}+\kappa\right)^2}\left\{\lambda_{1,t}\gamma_t x_{t-1}(\beta b_{t-1}^{\pi}+\kappa)\right. \\ & \left.+E_t[\lambda_{1,t+1}\beta\gamma_{t+1}((\beta b_t^{\pi}+\kappa)x_{t+1}+u_{t+1}-b_t^{\pi}2x_t)]\right\}\end{aligned} \tag{45}$$

Plugging (43) into equation (29), we get the following law of motion of b^{π} along any optimal path:

$$b_t^{\pi}=b_{t-1}^{\pi}+\gamma_t x_{t-1}\left[(\beta b_{t-1}^{\pi}+\kappa)\Phi_1\left(b_{t-1}^{\pi}\right)u_t+u_t-x_{t-1}b_{t-1}^{\pi}\right]+\gamma_t x_{t-1}(\beta b_{t-1}^{\pi}+\kappa)\Phi_2\left(s_t\right) \tag{46}$$

Using analogous arguments, we get that:

$$b_t^{x}=b_{t-1}^{x}+\gamma_t x_{t-1}\left[\Phi_1\left(b_{t-1}^{\pi}\right)u_t-x_{t-1}b_{t-1}^{x}\right]+\gamma_t x_{t-1}\Phi_2\left(s_t\right) \tag{47}$$

Our aim is to rewrite equations (46)-(47) as a Stochastic Recursive Algorithm (SRA hereafter) in a form that can be analyzed using the stochastic approximation tools. To do so, we start defining the vector of the state variables of the algorithm $Y_t \equiv [x_t, x_{t-1}, u_t, \gamma_t,]'$.[18] Hence, we can rewrite (46)-(47) as follows:

$$\begin{aligned} b_t^{\pi} &= b_{t-1}^{\pi}+\gamma_t\mathcal{H}_{\pi}\left(b_{t-1}^{\pi},Y_t^2,Y_t^3\right)+\gamma_t^2\rho_{\pi}\left(b_{t-1}^{\pi},b_{t-1}^{x},Y_t^2,Y_t^3,Y_t^4\right) \\ b_t^{x} &= b_{t-1}^{x}+\gamma_t\mathcal{H}_{x}\left(b_{t-1}^{\pi},Y_t^2,Y_t^3\right)+\gamma_t^2\rho_{x}\left(b_{t-1}^{\pi},b_{t-1}^{x},Y_t^2,Y_t^3,Y_t^4\right)\end{aligned}$$

where Y_t^i denotes the i-th entry of the Y_t vector, and:

$$\begin{aligned}\mathcal{H}_{\pi}\left(b_{t-1}^{\pi},Y_t^2,Y_t^3\right) &\equiv x_{t-1}\left[(\beta b_{t-1}^{\pi}+\kappa)\Phi_1\left(b_{t-1}^{\pi}\right)u_t+u_t-x_{t-1}b_{t-1}^{\pi}\right] \\ \mathcal{H}_{x}\left(b_{t-1}^{\pi},Y_t^2,Y_t^3\right) &\equiv x_{t-1}\left[\Phi_1\left(b_{t-1}^{\pi}\right)u_t-x_{t-1}b_{t-1}^{x}\right] \\ \rho_{\pi}\left(b_{t-1}^{\pi},b_{t-1}^{x},Y_t^2,Y_t^3,Y_t^4\right) &\equiv x_{t-1}(\beta b_{t-1}^{\pi}+\kappa)\frac{\Phi_2\left(s_t\right)}{\gamma_t} \\ \rho_{x}\left(b_{t-1}^{\pi},b_{t-1}^{x},Y_t^2,Y_t^3,Y_t^4\right) &\equiv x_{t-1}\frac{\Phi_2\left(s_t\right)}{\gamma_t}\end{aligned}$$

[18]Note that the vector of state variables used for the convergence analysis is different from those used in the solution of the optimization problem.

If we define $\theta_t \equiv [b_t^\pi, b_t^x]'$, and:

$$\mathcal{H}(\cdot) \equiv \begin{pmatrix} \mathcal{H}_\pi(\cdot) \\ \mathcal{H}_x(\cdot) \end{pmatrix}, \quad \rho(\cdot) \equiv \begin{pmatrix} \rho_\pi(\cdot) \\ \rho_x(\cdot) \end{pmatrix}$$

equations (46)-(47) can be written as:

$$\theta_t = \theta_{t-1} + \gamma_t \mathcal{H}(\theta_{t-1}, Y_t) + \gamma_t^2 \rho(\theta_{t-1}, Y_t) \tag{48}$$

which is a SRA in the standard form studied in the Evans and Honkapohja (2001). To study the asymptotic behavior of θ_t, we analyze the solutions and stability of the Ordinary Differential Equation (ODE) associated to (48):

$$\frac{d\theta}{d\tau} = h(\theta) \equiv E\mathcal{H}\left(b^\pi, \widehat{Y}_t^2, \widehat{Y}_t^3\right) \tag{49}$$

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 $\theta_{t-1} = \theta$. 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.[19] Note that x_{t-1} does not depend on u_t; this implies that:

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

The only possible rest point of the ODE (49) is clearly $\theta = 0$. Moreover it is (locally) stable, since the Jacobian:

$$Dh(\theta) = \begin{pmatrix} -Ex_{t-1}^2(\theta) - b^\pi \frac{\partial Ex_{t-1}^2(\theta)}{\partial b^\pi} & -b^\pi \frac{\partial Ex_{t-1}^2(\theta)}{\partial b^x} \\ -b^x \frac{\partial Ex_{t-1}^2(\theta)}{\partial b^\pi} & -Ex_{t-1}^2(\theta) - b^x \frac{\partial Ex_{t-1}^2(\theta)}{\partial b^x} \end{pmatrix} \tag{50}$$

[19] The proof is available from the authors upon request.

has both eigenvalues smaller than zero when evaluated in $\theta = 0$.[20] 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 (50) we conclude that this E-stability result is independent of parameters' values.

Remark 1. *The Jakobian* (50) *has negative eigenvalues for any value of the structural parameters.*

Evans and Honkapohja (2001) derive an equivalence result between E-stability and convergence under learning. This theorem, which draws on arguments contained in Benveniste, Métivier, and Priouret (1990), cannot directly be applied to our problem, since the state variables' law of motion does not satisfy the required assumptions.[21] However, it turns out that we can adapt their arguments, and prove the following result.[22]

Proposition 3. *Let θ evolve according to* (48)*. If $\overline{\theta}$ is E-stable, then it is locally stable under adaptive learning.*[23]

Proof. See the Appendix. □

Proposition 3 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. Equations (18) and (19) then show that in the limit agents expect zero inflation and output-gap. Substituting this together with $\gamma_t \to 0$ into the FOC (41) and the PC (2) implies that both output and inflation converges to the IT equilibrium (14) (15).

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

[21] 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 in page 216 of Benveniste, Métivier, and Priouret (1990) is not satisfied (we cannot prove the existence of a solution to the Poisson equation).

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

[23] For an explicit definition of what "locally stable under adaptive learning" means, see Evans and Honkapohja (2001) page 275.

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

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

3 Policy implications

In the previous section we established that the optimal policy drives agents' beliefs to the inflation targeting equilibrium. In order to explain the intuition behind this result, in this section we describe the short and long run policy tradeoffs.

3.1 Welfare implications

In order to quantify the long run and short run tradeoffs, we use numerical methods. We use the FOCs (41)-(42) and solve for $\lambda_{1,t}$ and x_t, using a collocation algorithm. We approximate the control variables with Chebychev polynomials, as functions of the state variables (x_{t-1}, b^{π}_{t-1} and u_t)[24]. The optimal approximated policy functions are then used to simulate the series.

The benchmark calibration is taken from Woodford (1999) (see table 1). In order to avoid the effect of a changing gain parameter, and focus entirely on the short versus long run trade-off, we simulate the model for a small constant gain parameter. The reason is that, with decreasing gain learning, the first observations of the simulated series are strongly affected by the value of the gain parameter $\gamma_t = \frac{1}{t}$. Simulations starting from period 1, where $\gamma_1 = 1$, are quantitatively different from simulations starting from period 1000, where $\gamma_1 = 0.001$. To abstract from the effect of a changing gain parameter, we prefer to present our results only

[24]We make use of the Miranda-Fackler CompEcon Toolbox. We use tensor product to project the multimensional state space on the policy space, and Gaussian quadrature to compute the expectation operators. The solution is found by using a version of the Broyden algorithm for nonlinear equations coded by Michael Reiter. Uniqueness of the solution might be an issue, since the Kuhn-Tucker conditions are only necessary in our setup. However, we experimented with several initial conditions and different interpolation techniques, and the solution did not change.

Table 1: Parameters

Parameter	Value
β	0.99
σ	0.157
κ	0.024
α	0.04
γ	0.05
σ_u^2	0.07

Gaussian cost-push shock, $Eu = 0$.

for constant gain. However, the qualitative behaviour of the series is the same under constant and decreasing gain. The decreasing gain results are available upon request. We set $\gamma = 0.05$, which is a value consistent with estimates for the US economy (see Milani (2007), Branch and Evans (2006) and Slobodyan and Wouters (2012)). Robustness checks for several of the model parameters have been performed and are available upon request.

Figure 1: Dynamics of b^π and b^x under constant gain, benchmark parameterization, $\gamma = .05$

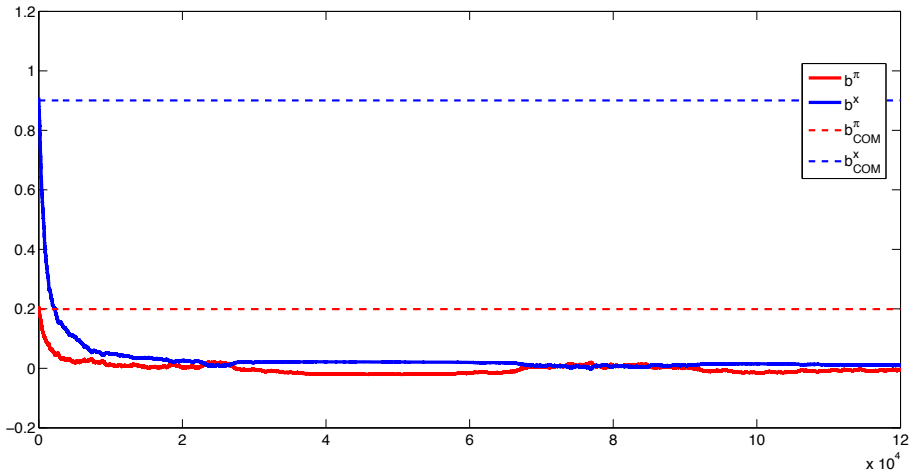


Figure 1 provides an example of the evolution of the learning coefficients b_t^π and b_t^x for a single simulated path of u: it shows that the optimal policy drives beliefs to the IT equilibrium, and this equilibrium is stable: once reached this

equilibrium, agents' beliefs remain very close to it. The learning coefficients b_t^π and b_t^x converge to zero, thus inflation and output gap expectations (18-19) do not depend on lagged output gap and converge to zero too.

The welfare benefits of the optimal policy are substantial. In table 2 we compare welfare losses from the optimal policy to the losses obtained with a Taylor-type rule that keeps learning agents in the PLT equilibrium (as shown in Evans and Honkapohja (2006)).[25] We perform a Monte Carlo with simulation length of 5000 periods and a cross-sectional sample size of 10000, and express welfare losses as consumption equivalent in terms of steady state consumption. The loss of PLT is 63% percent higher than that of the optimal rule, when starting from beliefs consistent with PLT. The same measure is 35% higher for the optimal policy when we start from beliefs consistent with IT. Therefore, our optimal policy is significantly better than a PLT rule.

Table 2: Consumption equivalents

Initial beliefs	Optimal policy	Price level targeting	Ratio ($\frac{PLT}{OP}$)
Inflation targeting	0.000744	0.001004	1.35
Price level targeting	0.000411	0.000673	1.63

$\gamma = 0.05$.

3.2 Welfare decomposition: short vs long run

The optimal policy's welfare gains (with respect to the PLT Taylor rule) can be decomposed in short run losses and long term benefits. To illustrate this, in Figure 2 we produce a consumption equivalent measure of welfare losses over a rolling window, for the optimal policy (blue line). We then compare our optimal policy with the Taylor rules that keep beliefs respectively to PLT (red line) and IT (black line) equilibria[26].

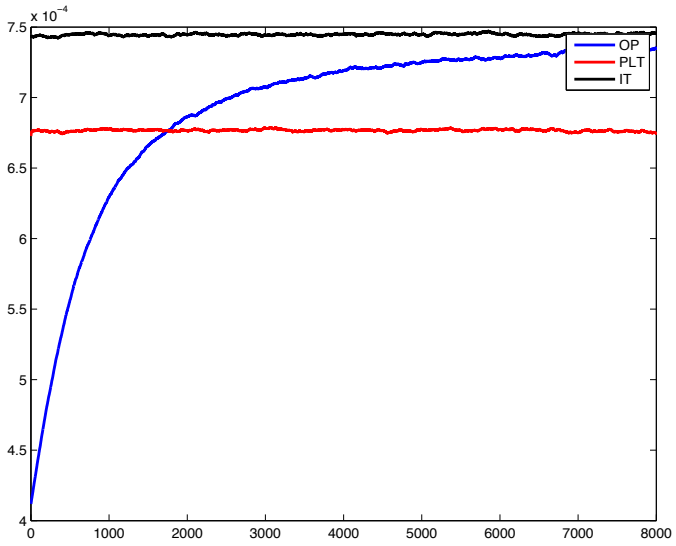
The graph should be read in the following way: a point at time t in the blue line corresponds to the consumption equivalent measure of the optimal policy start-

[25] This rule also guarantees determinacy under RE.

[26] The latter is taken from Evans and Honkapohja (2003).

ing from the belief in period t, averaged over 10000 simulations of 2000 periods' length. For example, at period 1000 we have the average consumption equivalent for the optimal policy starting with beliefs at period 1000. PLT and IT policies are obtained in an analogous way. The only difference is in the initial beliefs: for PLT and OP, we set them at the PLT value. IT policy instead is simulated starting from IT beliefs. In this way, we expect to see the optimal policy welfare measure converge to the IT one.

Figure 2: 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, $\gamma = 0.05$.

First of all, Figure 2 shows that the welfare ranking of the two Taylor rules under learning are similar to the one under RE: the per period welfare losses of IT are higher than that of PLT in each period. Moreover, from this graph, we can see that the policymaker sacrifices long run efficiency for short run gains. In the initial periods, welfare losses from the optimal policy are clearly smaller than the PLT rule. However, optimal policy's losses are larger than those generated by

a PLT rule in the long run, as the optimal policy tends asymptotically to the IT policy. The intuition is straightforward: in the short run, the optimal policy can respond quickly to a shock due also to almost constant expectations, while PLT policy response need to anchor future inflation expectations by committing to a long sequence of future output contractions. Hence, in the short run the optimal policy has an advantage in terms of welfare: it can close the wedge created by the shock in a very short time without large impact on inflation expectations. However, in the long run, as the optimal policy drives agents' beliefs away from PLT, the policymaker looses its ability to anchor agents' inflation expectations. Its policy therefore resembles more the IT policies, which are welfare inferior to PLT.

3.3 Short run policy incentives

The short run gains come from the well known time inconsistency problem of price level targeting and the sluggishness of agents' beliefs. The time inconsistency is standard: if given the chance, the central bank has an incentive to renege its commitments and choose a different policy which is optimal at the time the decision is taken. Under rational expectations, any deviation from a commitment will be immediately spotted by agents, making any future commitment of the central bank not credible anymore. However, under learning things are different: small deviations from PLT, for example, can be "interpreted" by agents as a mistake in their estimated model of the economy.

This can be easily illustrated by looking at the first order conditions of the central bank. Let us first assume that agents do not update their beliefs, so $\gamma_t = 0$ and learning coefficients are not updated. Let us also assume that by some period k agents believe that the central bank is implementing the PLT policy, hence $b_k^x = b^x$ and $b_k^\pi = b^\pi$. Agents' expectations are then equal to the RE equilibrium under PLT: $E_t^* \pi_{t+1} = b^\pi x_t, E_t^* x_{t+1} = b^x x_t$, $\forall t \geq k$. However, given these assumptions, the optimal policy is strikingly different from PLT. By combining the FOCs (41)-

(42) and the PC (2) we get

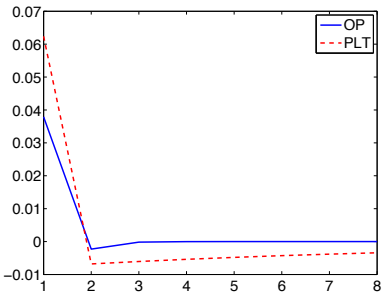
$$\begin{aligned} x_t &= -\frac{\beta b^\pi + \kappa}{\alpha + (\beta b^\pi + \kappa)^2} u_t \\ \pi_t &= \frac{\alpha}{\alpha + (\beta b^\pi + \kappa)^2} u_t. \end{aligned} \tag{51}$$

When the learning is shut down, the central bank optimal policy does not depend on x_{t-1}, as in the PLT equilibrium (7-8). Instead, its policy is qualitatively similar to the "leaning against the wind" strategy of IT: after a positive shock, the CB decreases current output gap in order to avoid a huge increase in current inflation. This reaction is stronger the larger is b^π, i.e. the further away expectations are from the IT policy: intuitively, the higher b^π, the stronger is the trade-off between inflation and output (from the Phillips curve (2)), and therefore the stronger is the incentive of the central bank to "fool" agents.

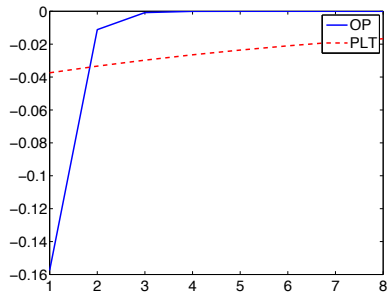
Things are slightly different if we allow for learning, i.e. if $\gamma_t > 0$. This implies that agents are learning from the realized allocations and they eventually understand when the central bank deviates from the PLT. Agents endowed with rational expectations would immediately loose any faith in the central bank credibility after a deviation, and they would assume that the prevailing equilibrium in the future will be IT. On the other hand, learning agents revise their beliefs in a more sluggish fashion. Because of this sluggishness, and similarly to the case in which there is no learning at all, the policymaker has still an incentive to "surprise" the households repeatedly, by choosing allocations different from expected ones. The central bank implements these surprises by aggressively contracting output to disinflate. Figures 3a and 3b show the impulse response functions for inflation and output, starting from PLT beliefs. Compared to PLT, after a positive cost-push shock the OP engineers a much bigger output contraction in order to keep inflationary pressures at bay. With respect to PLT, the welfare gain of lower inflation outweighs the welfare loss of a larger output gap, since in the New Keynesian model price rigidity is the most important friction. When inflation is lower, firms that cannot adjust their prices have a less distorted price, and their output is closer to the (flexible prices) efficient output. However, this policy is not consistent with agents' beliefs, since they were expecting a PLT policy, and hence

they will make forecasting mistakes. Households will learn from these mistakes by updating their beliefs as shown in figure 1. In the limit households learn that the central bank is not implementing a PLT policy, and their beliefs slowly converge to the IT equilibrium.

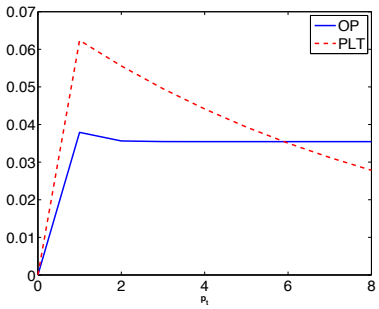
Figure 3: Impulse responses after a one standard deviation cost-push shock, under optimal policy under learning (OP) and price-level targeting policy (PLT) , starting with initial beliefs corresponding to the rational expectations PLT equilibrium, with $\gamma = 0.05$.



(a) Inflation π_t



(b) Output-gap x_t



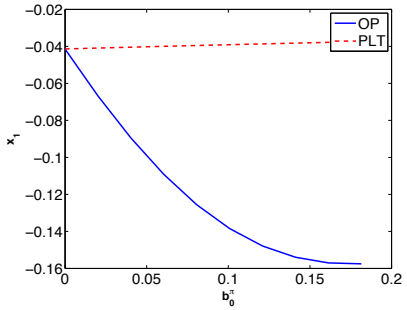
(c) Price level p_t

The impulse response of the price level shows even more strikingly how our optimal policy differs from PLT. Figure 3c shows that the optimal response after a positive cost-push shock is to allow the price level to raise permanently (similarly to what would happen under an IT rule), while under PLT the central bank would bring the price level back to the target. In other words, price level stabilization is not optimal when agents are learning, and the central bank should let prices absorb shocks in a permanent way.

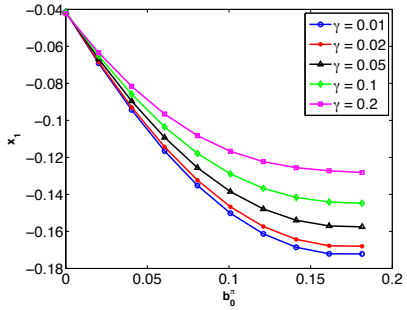
3.4 Non-linearity of the optimal policy

The optimal policy is a non-linear function of household expectations, therefore it cannot be expressed as a Taylor type rule. To illustrate this, Figure 4 (panel a) shows the first period output contraction x_1 that the central bank engineers after a positive cost-push shock, for different inflation beliefs b_0^π.

Figure 4: Impulse responses of inflation and output gap 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) 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

The PLT impact is a linear, slightly increasing function of the beliefs by construction, since this is a linear expectations-based interest rate rule derived in Evans and Honkapohja (2006). The further away beliefs are from the PLT equilibrium, the larger the output contraction that PLT policy engineers, in order to drive household beliefs back to the PLT equilibrium. Optimal policy, on the other hand, is non-linear in inflation beliefs and, in contrast to PLT, it is decreasing in b_0^π. In other words, the closer household beliefs are to the PLT equilibrium, the larger is the output contraction engineered, and the larger is the difference with PLT. The reason for this is that the closer beliefs are to PLT, the more the central bank can decrease next periods' inflation expectations by contracting output without substantially affecting agents' beliefs, therefore the incentives to exploit the inflation-output tradeoff are bigger.

Those same incentives also depend on the gain parameter. The smaller is the gain parameter, the more hawkish the central bank is: it engineers a bigger output contraction (Figure 4, panel b). With a small gain parameter agents learn slowly and optimal policy can exploit the Phillips curve and aggressively disinflate without loosing its ability to affect private sectors' expectations.

4 Discussion

The results we derived show that optimality of price level targeting is not robust to relaxing rationality bounds of private agents. When private agents use past data to form beliefs about the future instead of being fully rational, price level targeting is still beneficial in effectively anchoring inflation expectations, but monetary policy has strong short run incentives to deviate from it. These incentives arise because learning agents need time to uncover that the central bank has deviated from PT, and in the meantime the policymaker can exploit the inflation-output tradeoff and disinflate by aggressively contracting output. This policy comes at a cost, private agents eventually gather enough data and understand that the central bank is deviating from PT. The economy converges to IT and the central bank looses its ability to anchor private expectations. We show that the short run gains of this policy outweigh long run losses, therefore it is optimal for the central bank to succumb to the temptation and deviate from price level targeting.

The central bank incentives that arise in our framework have been previously ignored by proponents of price level targeting under learning (see Evans and Honkapohja (2006), Aoki and Nikolov (2006), Gaspar, Smets, and Vestin (2007)). These authors showed that price level targeting is a learnable equilibrium: if expectations are perturbed out of the price level targeting equilibrium, the central bank can implement a policy that makes agents learn the price level targeting equilibrium again. However, once central bank 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 learnability of an equilibria, as it is traditionally done in the

literature (see Evans and Honkapohja (2001)). Even a learnable equilibria might not arise when interactions between agents are taken into account. The incentives of a rational player (in our model the central bank) depend on what type of other player she interacts with. Adaptive players are different from rational players even after they learned a rational expectations equilibria, and their forecast could not be distinguished from that of a rational agent. One difference is the speed of revising beliefs. A rational agent would immediately understand if the central bank has deviated from PT and would immediately switch to the IT equilibrium. A learning agent on the other hand needs time to gather sufficient amount of data to understand that the central bank deviated from PT. 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 PT (see Kurozumi (2008)). For learning agents, on the other hand, off-equilibrium threats are not possible, since they simply form beliefs based on realized outcomes. A rational opponent to learning agents takes this into account and chooses her strategy accordingly.

In our setup, the central bank incentives to deviate from PT cannot be turned around by appointing a conservative central banker, in a way analogous to what suggested in Rogoff (1985). Even if the central banker cares strongly about dampening inflation fluctuations, these incentives remain.

To be clear, we are not claiming that price level targeting should never be used by central banks. Our claim is that central banks have incentives to deviate from this policy when agents are learning. As learning is found to be empirically relevant,[27] we think these central bank incentives should not be neglected. One caveat of our findings is that many empirical and policy-oriented models are a more complex representation of the economy than our setup (see Smets and Wouters (2007)). For example, many researchers add various exogenous sources of persistence. We cannot exclude that other sources of persistence might interact with learning and provide different policy incentives than in our paper. Unfortunately, introducing these features in our model would make it analytically and numerically intractable.

[27]See for example Del Negro and Eusepi (2011), Slobodyan and Wouters (2012), Molnar and Ormeno (2014).

A few of our assumptions play an important role in our findings, and therefore we would like to discuss their limitations. Firstly, we conduct our analysis by assuming a specific learning algorithm. Even if this learning algorithm is widely used in the literature, and it is consistent with the rational expectation equilibrium, it might seem arbitrary. Yet it is equally arbitrary to assume that there is no learning component to private expectations, especially since this is at odds with recent empirical findings. Ultimately, how people form expectations is an empirical issue, which is yet unsettled. We agree with Marcet and Nicolini (2003) that rationality bounds should be placed on learning, and we think it would be beneficial to expand both empirical and theoretical research in this direction. As this is beyond the scope of this paper, for the time being, we would like to emphasize that it should not be ignored that central bank incentives change when there is a learning component to agents' beliefs.

Secondly, we assume the central bank knows the exact learning algorithm agents use. This is undoubtedly a strong assumption, nevertheless an analogously strong assumption is regularly made in optimal policy research with rational agents, where the policymaker knows that agents are rational. We do think it is worth making our extreme assumption in order to understand optimal policy under the polar case of adaptive learning. Assessing the consequences of intermediate forms of rationality would require their explicit modelling, and we do not rule out the possibility that different policy incentives would arise in an alternative setting.

5 Conclusions

This paper has shown that stabilizing prices is a bad strategy if agents do not have perfectly rational expectations. We have examined an optimal monetary policy problem in a setup where agents are adaptive learners, and our main result shows that a benevolent central bank should not counteract the effects of economic fluctuations on the price level. The optimal policy instead let prices absorb the effects of shocks in a permanent way. Qualitatively, the optimal policy resembles a "leaning against the wind" policy.

There is a large literature that examines the policy implications of standard monetary policy prescriptions in economies where agents are adaptive learners. Contrary to the approach in the previous literature, our optimal policy design takes into account the incentives of the policymaker. In this setup, a good policymaker sacrifices the long run benefits of stabilizing prices for the short run gains coming from exploiting the inflation-output trade-off, by taking into account the expectations formation mechanism of the private sector. This result is in line with some observed features of the practice of monetary policy. Central banks routinely monitor private sector's expectations, and are reluctant to introduce price level stabilization as their official policy objective.

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 central bank change with the expectation formation mechanism of the private sector, and policy prescriptions derived without acknowledging this fact can be misleading.

An important question is how general our result is. We conjecture that it is common to a large class of Stackelberg games, where a leader makes optimal decisions by explicitly taking into account the expectations formation mechanism of the follower. We leave this extension to future research.

References

Adam, K. (2003): "Learning and Equilibrium Selection in a Monetary Overlapping Generations Model with Sticky Prices," *Review of Economic Studies*, 70, 887–908.

Adam, K., and A. Marcet (2011): "Internal rationality, imperfect market knowledge and asset prices," *Journal of Economic Theory*, 146(3), 1224–1252.

Ambler, S. (2009): "Price-Level Targeting And Stabilisation Policy: A Survey," *Journal of Economic Surveys*, 23(5), 974–997.

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

Beggs, A. (2005): "On the convergence of reinforcement learning," *Journal of Economic Theory*, 122(1), 1 – 36.

Benveniste, A., M. Métivier, and P. Priouret (1990): *Adaptive Algorithms and Stochastic Approximations*. Berlin: Springer-Verlag.

Borgers, T., and R. Sarin (1997): "Learning Through Reinforcement and Replicator Dynamics," *Journal of Economic Theory*, 77(1), 1–14.

Branch, W., and B. McGough (2011): "Business cycle amplification with heterogeneous expectations," *Economic Theory*, 47(2), 395–421.

Branch, W. A., and G. W. Evans (2006): "A simple recursive forecasting model," *Economics Letters*, 91(2), 158–166.

Brock, W. A., and C. H. Hommes (1997): "A Rational Route to Randomness," *Econometrica*, 65(5), 1059–1096.

Bullard, J., G. W. Evans, and S. Honkapohja (2008): "Monetary Policy, Judgment and Near-Rational Exuberance," *American Economic Review*, 98, 1163–1177.

Bullard, J., and K. Mitra (2002): "Learning about Monetary Policy Rules," *Journal of Monetary Economics*, 49(6), 1105–1129.

Calvo, G. (1983): "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, 12(3), 383–398.

Clarida, R., J. Gali, and M. Gertler (1999): "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37(2), 1661–1707.

Del Negro, M., and S. Eusepi (2011): "Fitting observed inflation expectations," *Journal of Economic Dynamics and Control*, 35(12), 2105–2131.

Eggertsson, G. B., and M. Woodford (2003): "The Zero Bound on Interest Rates and Optimal Monetary Policy," *Brookings Papers on Economic Activity*, 34(1), 139–235.

Eusepi, S., and B. Preston (2010): "Central Bank Communication and Expectations Stabilization," *American Economic Journal: Macroeconomics*, 2(3), 235–71.

Evans, G. W., and S. Honkapohja (1998): "Convergence of learning algorithms without a projection facility," *Journal of Mathematical Economics*, pp. 59–86.

——— (2001): *Learning and Expectations in Macroeconomics*. Princeton: Princeton University Press.

——— (2003): "Expectations and the Stability Problem for Optimal Monetary Policies," *Review of Economic Studies*, 70(4), 807–824, available at http://ideas.repec.org/a/bla/restud/v70y2003i4p807-824.html.

——— (2006): "Monetary Policy, Expectations and Commitment," *The Scandinavian Journal of Economics*, 108(1), 15–38.

Fudenberg, D., and D. K. Levine (1998): *The Theory of Learning in Games*, vol. 1 of *MIT Press Books*. The MIT Press.

Gali, J. (2003): "New Perspectives on Monetary Policy, Inflation, and the Business Cycle," in *Advances in Economic Theory*, ed. by M. Dewatripont, L. Hansen, and S.Turnovsky. Cambridge: Cambridge University Press.

Gaspar, V., F. Smets, and D. Vestin (2006): "Optimal Monetary Policy under Adaptive Learning," Computing in Economics and Finance 2006 183, Society for Computational Economics.

Gaspar, V., F. Smets, and D. Vestin (2007): "Is time ripe for price level path stability?," Working Paper Series 0818, European Central Bank.

Gorodnichenko, Y., and M. D. Shapiro (2007): "Monetary policy when potential output is uncertain: Understanding the growth gamble of the 1990s," *Journal of Monetary Economics*, 54(4), 1132–1162.

Jaskiewicz, A., and A. S. Nowak (2011): "Stochastic Games with Unbounded Payoffs: Applications to Robust Control in Economics," *Dynamic Games and Applications*, 1, 253–279.

Kurozumi, T. (2008): "Optimal sustainable monetary policy," *Journal of Monetary Economics*, 55(7), 1277–1289.

Marcet, A., and J. P. Nicolini (2003): "Recurrent Hyperinflations and Learning," *American Economic Review*, 93(5), 1476–1498.

Marcet, A., and T. J. Sargent (1989a): "Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information," *Journal of Political Economy*, 97(6), 1306–1322.

——— (1989b): "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," *Journal of Economic Theory*, 48(2), 337–368.

Marimon, R., and S. Sunder (1993): "Indeterminacy of Equilibria in a Hyperinflationary World: Experimental Evidence," *Econometrica*, 61(5), 1073–107, available at http://ideas.repec.org/a/ecm/emetrp/v61y1993i5p1073-107.html.

Meh, C. A., J.-V. Ros-Rull, and Y. Terajima (2010): "Aggregate and welfare effects of redistribution of wealth under inflation and price-level targeting," *Journal of Monetary Economics*, 57(6), 637–652.

Milani, F. (2006): "A Bayesian DSGE Model with Infinite-Horizon Learning: Do "Mechanical" Sources of Persistence Become Superfluous?," *International Journal of Central Banking*, 2(3), 87–106.

——— (2007): "Expectations, Learning and Macroeconomic Persistence," *Journal of Monetary Economics*, 54(7), 2065–2082.

Molnar, K. (2007): "Learning with Expert Advice," *Journal of the European Economic Association*, 5(2-3), 420–432.

Molnar, K., and A. Ormeno (2014): "Using Survey Data of Inflation Expectations in the Estimation of Learning and Rational Expectations Models," *Journal of Money Credit and Banking*, accepted.

Molnar, K., and S. Santoro (2014): "Optimal Monetary Policy when Agents are Learning," *European Economic review*, 66, 39–62.

Rogoff, K. (1985): "The Optimal Degree of Commitment to an Intermediate Monetary Target," *The Quarterly Journal of Economics*, 100(4), 1169–89.

Rotemberg, J. J., and M. Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in *NBER Macroeconomics Annual 12*, ed. by B. Bernanke, and J. J. Rotemberg. Cambridge (MA): MIT Press.

Slobodyan, S., and R. Wouters (2012): "Learning in an estimated medium-scale DSGE model," *Journal of Economic Dynamics and Control*, 36(1), 26–46.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97(3), 586–606.

Wolman, A. L. (2005): "Real Implications of the Zero Bound on Nominal Interest Rates," *Journal of Money, Credit and Banking*, 37(2), 273–96.

Woodford, M. (1999): "Optimal Monetary Policy Inertia," *The Manchester School, Supplement*, 67(0), 1–35.

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

Yun, T. (1996): "Nominal price rigidity, money supply endogeneity, and business cycles," *Journal of Monetary Economics*, 37(2-3), 345–370, available at http://ideas.repec.org/a/eee/moneco/v37y1996i2-3p345-370.html.

Appendix

In this Appendix we prove proposition 3. To do so, we first show a series of intermediate results.

First of all, we state and prove the following technical Lemma.

Lemma 3. *Let $\lambda_{1,t}$ be a stationary solution of (42), and suppose that θ_t is fixed at some θ; then, for any compact $Q \subset \mathbb{R}^2$, there exists a positive constant C_λ such that:*

$$|\lambda_{1,t}| \leq C_\lambda \left(1 + |u_t|^2\right) \tag{A.1}$$

for any $\theta \in Q$.

Proof. Solving forward equation (42), we get that any stationary solution must satisfy:

$$\begin{aligned} \lambda_{1,t} &= \beta^2 E_t \sum_{i=1}^{\infty} \left\{ \beta^i \left[\left((\beta b^\pi + \kappa) x_{t+1+i} + u_{t+1+i} \right) x_{t+1+i} \right] \Pi_{j=0}^{i} \vartheta_{t+j} \right\} + \\ &\quad + \beta^2 E_t \left[\left((\beta b^\pi + \kappa) x_{t+1} + u_{t+1} \right) x_{t+1} \right] \end{aligned} \tag{A.2}$$

where ϑ_{t+j} is defined as follows:

$$\vartheta_t = 1, \;\; \vartheta_{t+j} = 1 - \gamma_{t+j} x_{t+j-1} \left(x_{t+j-1} - \beta x_{t+j} \right) \;\; \text{for } j > 0$$

Let $\overline{x}\left(u_t\right)$ be defined as in the statement of Proposition 2, let:

$$\overline{\pi}\left(u_t\right) \equiv M_Q \overline{x}\left(u_t\right) + u_t$$

where $M_Q \equiv \max_{\theta \in Q} (\beta b^\pi + \kappa)$.[28] Moreover, note that for any $j > 0$:

$$\begin{aligned} |\vartheta_{t+j}| &= \left| 1 - \gamma_{t+j} x_{t+j-1} \left(x_{t+j-1} - \beta x_{t+j} \right) \right| \leq 1 + \gamma_{t+j} \left| x_{t+j-1} \right|^2 + \beta \gamma_{t+j} \left| x_{t+j-1} x_{t+j} \right| \\ &< 1 + \gamma_{1+j} \left| \overline{x}\left(u_{t+j-1}\right) \right|^2 + \beta \gamma_{1+j} \left| \overline{x}\left(u_{t+j-1}\right) \overline{x}\left(u_{t+j}\right) \right| \equiv \overline{\vartheta}_{t+j} \end{aligned}$$

where we used the triangle inequality, the fact that the sequence of gains is decreasing, and the result of Proposition 2 that at an optimum we must have $|x_t| < \overline{x}\left(u_t\right)$.

[28] This maximum exists, since the function is continuous and Q is compact by assumption.

Because the stochastic process of u is assumed to be iid, it follows that $\overline{\vartheta}_{t+j}$ is independent of $\overline{x}(u_{t+1+i})$ and $\overline{\pi}(u_{t+1+i})$, for any $j \leq i$. Using this observation, the bounds derived on x, $((\beta b^{\pi} + \kappa)x + u)$ and ϑ, the triangle inequality, the Schwartz inequality, the monotonicity of the expectation operator, we can write:

$$|\lambda_{1,t}| \leq \beta^2 M_{x,\pi} E_t \sum_{i=1}^{\infty} \left\{ \beta^i \Pi_{j=0}^{i} \overline{\vartheta}_{t+j} \right\} + \beta^2 M_{x,\pi}$$

where $M_{x,\pi} \equiv E_t \overline{x}(u_{t+1+i})\,\overline{\pi}(u_{t+1+i})$ which is constant for any t and any i because of the iid assumption. Note that the series in the RHS of the above inequality converges, since $\beta < 1$ and $\lim_{j\to\infty} E_t \overline{\vartheta}_{t+j} = 1$. Finally, note that the only $\overline{\vartheta}_{t+j}$ that depends on u_t is $\overline{\vartheta}_{t+1}$; hence, we can write the above inequality as follows:

$$\begin{aligned}
|\lambda_{1,t}| \;\leq\; & \beta^2 M_{x,\pi} E_t \sum_{i=2}^{\infty} \left\{ \beta^i \left[\Pi_{j=2}^{i} \overline{\vartheta}_{t+j} \right] \left[1 + \gamma_2 \left| \overline{x}(u_t) \right|^2 + \beta\gamma_2 \left| \overline{x}(u_t)\,\overline{x}(u_{t+1}) \right| \right] \right\} + \\
& \beta^3 M_{x,\pi} E_t \left[1 + \gamma_2 \left| \overline{x}(u_t) \right|^2 + \beta\gamma_2 \left| \overline{x}(u_t)\,\overline{x}(u_{t+1}) \right| \right] + \beta^2 M_{x,\pi} \\
=\; & \beta^2 M_{x,\pi} \left[1 + \gamma_2 \left| \overline{x}(u_t) \right|^2 \right] E_t \sum_{i=2}^{\infty} \beta^i \left[\Pi_{j=2}^{i} \overline{\vartheta}_{t+j} \right] + \\
& \beta^3 M_{x,\pi} \gamma_2 \overline{x}(u_t)\, E_t \sum_{i=2}^{\infty} \beta^i \left\{ \left[\Pi_{j=2}^{i} \overline{\vartheta}_{t+j} \right] \overline{x}(u_{t+1}) \right\} + \\
& \beta^3 M_{x,\pi} \left[1 + \gamma_2 \left| \overline{x}(u_t) \right|^2 + \beta\gamma_2 \overline{x}(u_t)\, E_t \overline{x}(u_{t+1}) \right] + \beta^2 M_{x,\pi} \\
\leq\; & \widehat{C}_{\lambda} \left(1 + |u_t| + |u_t|^2 \right) \qquad \text{(A.3)}
\end{aligned}$$

where we used the fact that, due to the iid assumption on u, the conditional expectations of the random variables considered in (A.3) are independent of t, and the definition of $\overline{x}(u_t)$ to get:

$$\overline{x}(s) = \epsilon \sqrt{\frac{(1-\beta)\,u^2 + \beta\sigma_u^2}{\alpha(1-\beta)}} \leq \epsilon \sqrt{\frac{\left[(1-\beta)\,|u| + \beta\sigma_u\right]^2}{\alpha(1-\beta)}} = \epsilon \frac{(1-\beta)\,|u|}{\sqrt{\alpha(1-\beta)}} + \epsilon \frac{\beta\sigma_u}{\sqrt{\alpha(1-\beta)}}$$

Finally, note that inequality (A.3) implies that there exists a C_{λ} such that (A.1) holds.[29] This completes the proof. □

[29] For example, $C_{\lambda} = 3\widehat{C}_{\lambda}$ would work.

We can now state and prove the following Proposition.

Proposition 4. *Let θ_t evolve according to (48), and fix an open set $D \subset \mathbb{R}^2$ around the point $\theta = 0$. Then, for any compact $Q \subset D$, there exist C and q such that for any $\theta \in Q$:*

$$|\rho(\theta, Y)| \leq C\left(1 + |Y|^q\right) \tag{A.4}$$

Proof. In what follows, we show that a bound of the form reported in the above inequality holds for the absolute value of any of the two components of the function $\rho(\cdot)$, which clearly implies (A.4).

Let's start from $\rho_\pi(\cdot)$; plugging equation (45) into the definition of this function we get:

$$\begin{aligned}
\left|\rho_\pi\left(b^\pi_{t-1}, b^x_{t-1}, Y^2_t, Y^3_t, Y^4_t\right)\right| &= \left| -x_{t-1}\frac{(\beta b^\pi_{t-1} + \kappa)}{\alpha + \left(\beta b^\pi_{t-1} + \kappa\right)^2}\left\{\lambda_{1,t}x_{t-1}(\beta b^\pi_{t-1} + \kappa)\right.\right. \\
&\quad \left.\left. + \beta\frac{\gamma_{t+1}}{\gamma_t}E_t[\lambda_{1,t+1}((\beta b^\pi_t + \kappa)x_{t+1} + u_{t+1} - b^\pi_t 2x_t)]\right\}\right| \\
&\leq \beta M_2 \left|E_t[\lambda_{1,t+1}((\beta b^\pi_t + \kappa)x_{t+1} + u_{t+1} - b^\pi_t 2x_t)]\right| \\
&\quad + M_1 \left|x_{t-1}\right|^2 \left|\lambda_{1,t}\right| \qquad \text{(A.5)}
\end{aligned}$$

where we used the triangle inequality and the fact that $\frac{\gamma_{t+1}}{\gamma_t} < 1$, and where:

$$M_1 \equiv \max_{\theta \in Q} \frac{(\beta b^\pi_{t-1} + \kappa)^2}{\alpha + \left(\beta b^\pi_{t-1} + \kappa\right)^2}, \quad M_2 \equiv \max_{\theta \in Q} \frac{(\beta b^\pi_{t-1} + \kappa)}{\alpha + \left(\beta b^\pi_{t-1} + \kappa\right)^2}$$

Using Lemma 3, we can write:

$$M_1 \left|x_{t-1}\right|^2 \left|\lambda_{1,t}\right| \leq M_1 \left|x_{t-1}\right|^2 C_\lambda \left(1 + |u_t|^2\right) \leq M_1 C_\lambda \left|x_{t-1}\right|^2 + 2M_1 C_\lambda \max\left\{\left|x_{t-1}\right|^2, |u_t|^2\right\}$$

Remember that the *max* between two real numbers define a norm on $\mathbb{R}^2$; by the well-known result that in a finite-dimensional normed linear space any two norms are equivalent, there exists a positive constant $\widehat{C}$ such that $\max\{z_1, z_2\} \leq \widehat{C}\left(|z_1| + |z_2|\right)$ for any $(z_1, z_2) \in \mathbb{R}^2$, where $|z_1| + |z_2|$ is a p-norm with $p = 1$. Hence,

we get:

$$\begin{aligned} M_1 C_\lambda \left|x_{t-1}\right|^2 + 2M_1 C_\lambda \max\left\{\left|x_{t-1}\right|^2, \left|u_t\right|^2\right\} &\leq M_1 C_\lambda \left|x_{t-1}\right|^2 + C_1\left(1 + \left|x_{t-1}\right|^2 + \left|u_t\right|^2\right) \\ &\leq C\left(1 + \left|x_{t-1}\right|^2 + \left|u_t\right|^2 + \left|\gamma_t\right|^2\right) \end{aligned}$$

Using similar arguments, we can obtain similar bounds for the term:

$$\beta M_2 \left|E_t[\lambda_{1,t+1}((\beta b_t^\pi + \kappa)x_{t+1} + u_{t+1} - b_t^\pi 2x_t)]\right|$$

which implies that the condition in the statement of the Proposition holds for $\rho_\pi(\cdot)$ with $q = 2$. In the case of $\rho_x(\cdot)$ the proof is analogous. □

The above Proposition implies that the assumptions made in Benveniste, Métivier, and Priouret (1990) on the SRA are satisfied by our model. In what follows, we show that the result that E-stability implies learnability holds even if we do not invoke their assumptions on the state variables' law of motion.

Following the steps described in Benveniste, Métivier, and Priouret (1990), Chapter 1 Part II, we rewrite the learning algorithm as follows

$$\theta_t = \theta_{t-1} + \gamma_t h\left(\theta_{t-1}\right) + \epsilon_{t-1} \tag{A.6}$$

where:

$$\epsilon_{t-1} = \gamma_t\left[\mathcal{H}\left(\theta_{t-1}, Y_t\right) - h\left(\theta_{t-1}\right) + \gamma_t \rho\left(\theta_{t-1}, Y_t\right)\right] \tag{A.7}$$

Heuristically, what we want to obtain are bounds on the fluctuations of the error term ϵ_{t-1}; more generally, we look for upper bounds of the expressions:

$$\epsilon_{t-1}\left(\phi\right) = \phi\left(\theta_t\right) - \phi\left(\theta_{t-1}\right) - \gamma_t \phi'\left(\theta_{t-1}\right) h\left(\theta_{t-1}\right) \tag{A.8}$$

where ϕ is an arbitrary twice continuously differentiable function from $\mathbb{R}^2$ to $\mathbb{R}$ with bounded second derivatives, and ϕ' is its gradient. In what follows we show that, fixing a compact set $Q \subset \mathbb{R}^2$, for any integer m there is a mean squares upper bound for the fluctuation:

$$\sup_{n \leq\ m \wedge \tau} \left|\sum_{k=0}^{n-1} \epsilon_k\left(\phi\right)\right| \tag{A.9}$$

where τ is the stopping time at which the process θ leaves for the first time the compact set Q:

$$\tau(Q) = \inf\{t : \theta_t \notin Q\} \tag{A.10}$$

Note that the assumptions on the function ϕ imply that:

$$\phi(\theta_{k+1}) - \phi(\theta_k) - (\theta_{k+1} - \theta_k)\phi'(\theta_k) = R(\phi, \theta_k, \theta_{k+1}) \tag{A.11}$$

where the function R, for all θ_k and θ_{k+1} has the upper bound[30]

$$|R(\phi, \theta_k, \theta_{k+1})| \leq |\theta_k - \theta_{k+1}|^2 \tag{A.12}$$

In order to find bounds on the error term $\epsilon_k(\phi)$, we can use equation (A.11) to decompose it as follows:

$$\begin{aligned}
\epsilon_k(\phi) &= \phi(\theta_{k+1}) - \phi(\theta_k) - \gamma_{k+1}\phi'(\theta_k)h(\theta_k) \\
&= \gamma_{k+1}\phi'(\theta_k)(\mathcal{H}(\theta_k, Y_{k+1}) - h(\theta_k)) + \gamma_{k+1}^2\rho(\theta_k, Y_{k+1}) + R(\phi, \theta_k, \theta_{k+1}) \\
&= \gamma_{k+1}\phi'(\theta_k)\left(\mathcal{H}(\theta_k, Y_{k+1}) + x_{k+1}^2\theta_k\right) + \gamma_{k+1}\phi'(\theta_k)\left(-h(\theta_k) - x_{k+1}^2\theta_k\right) + \\
&\quad \gamma_{k+1}^2\rho(\theta_k, Y_{k+1}) + R(\phi, \theta_k, \theta_{k+1})
\end{aligned} \tag{A.13}$$

Then, the running sum from $r < n$ to n of $\epsilon_k(\phi)$ on $\{n \leq \tau\}$ can be written as:

$$\sum_{k=r}^{n-1}\epsilon_k(\phi) = \sum_{k=r}^{n-1}\epsilon_k^1(\phi) + \sum_{k=r+1}^{n-1}\epsilon_k^2(\phi) + \sum_{k=r+1}^{n-1}\epsilon_k^3(\phi) + \sum_{k=r}^{n-1}\epsilon_k^4(\phi) + \sum_{k=r}^{n-1}\epsilon_k^5(\phi) + \sum_{k=r}^{n-1}\epsilon_k^6(\phi) + \eta_{n,r}(\phi) \tag{A.14}$$

[30] For all the details, see Benveniste, Métivier, and Priouret (1990) page 221.

where:

$$\epsilon_k^{(1)}(\phi) \equiv \gamma_{k+1}\phi'(\theta_k)\, x_k \left[(\beta b_k^\pi + \kappa)\Phi_1(b_k^\pi)\, u_{k+1} + u_{k+1}, \Phi_1(b_k^\pi)\, u_{k+1}\right]' \quad \text{(A.15)}$$

$$\epsilon_k^{(2)}(\phi) \equiv \gamma_{k+1}\phi'(\theta_k)\, x_k^2 (\theta_k - \theta_{k-1}) \quad \text{(A.16)}$$

$$\epsilon_k^{(3)}(\phi) \equiv (\gamma_k - \gamma_{k+1})\, \phi'(\theta_{k-1})\, x_k^2 \theta_{k-1} \quad \text{(A.17)}$$

$$\epsilon_k^{(4)}(\phi) \equiv \gamma_{k+1}\phi'(\theta_k)\, \theta_k \Phi_1^2(b_k^\pi)\left(\sigma_u^2 - u_{k+1}^2\right) \quad \text{(A.18)}$$

$$\epsilon_k^{(5)}(\phi) \equiv -\gamma_{k+1}\phi'(\theta_k)\, \theta_k \left(\Phi_2^2(s_{k+1}) + 2\Phi_2(s_{k+1})\, \Phi_1(b_k^\pi)\, u_{k+1}\right) \quad \text{(A.19)}$$

$$\epsilon_k^{(6)}(\phi) \equiv \gamma_{k+1}^2 \rho(\theta_k, Y_{k+1}) + R(\phi, \theta_k, \theta_{k+1}) \quad \text{(A.20)}$$

$$\eta_{n,r}(\phi) \equiv -\gamma_{r+1}\phi'(\theta_r)\, x_r^2 \theta_r + \gamma_n \phi'(\theta_{n-1})\, x_n^2 \theta_{n-1} \quad \text{(A.21)}$$

In the above decomposition we used the definition of $\mathcal{H}$ and the fact that in the optimum the square of the output gap is given by:

$$x_k^2 = \Phi_1^2\left(b_{k-1}^\pi\right) u_k^2 + \Phi_2^2(s_k) + 2\Phi_1\left(b_{k-1}^\pi\right) u_k \Phi_2(s_k)$$

The terms $\epsilon_k^{(2)}(\phi)$, $\epsilon_k^{(3)}(\phi)$, $\epsilon_k^{(6)}(\phi)$ and $\eta_{n,r}(\phi)$ are particular cases of expressions studied in Benveniste, Métivier, and Priouret (1990).[31] Hence, we concentrate on $\epsilon_k^{(1)}(\phi)$, $\epsilon_k^{(4)}(\phi)$ and $\epsilon_k^{(5)}(\phi)$. We start with $\epsilon_k^{(1)}(\phi)$.

Lemma 4. *There exist constants A_1 and q_1 such that:*

$$E_{y,a}\left\{\sup_{n\leq m} I(n \leq \tau)\left|\sum_{k=0}^{n-1} \epsilon_k^1(\phi)\right|\right\}^2 \leq A_1\left(1 + |y|^{q_1}\right)\sum_{k=0}^{m-1}\gamma_{k+1}^2 \quad \text{(A.22)}$$

where $E_{y,a}$ denotes expectations taken with respect to the distribution of histories induced by the transition probability of the Markov chain (Y_k, θ_k) with initial conditions $Y_0 = y$ and $\theta_0 = a$. Moreover, on $\{\tau \leq \infty\}$, $\sum_{k=0}^{n-1}\epsilon_k^1$ converges a.s. and in L^2.

Proof. Let's define:

$$\begin{pmatrix} (\beta b_k^\pi + \kappa)\Phi_1(b_k^\pi)\, u_{k+1} + u_{k+1} \\ \Phi_1(b_k^\pi)\, u_{k+1} \end{pmatrix} I(k+1 \leq \tau) \equiv \overline{Z}_k \quad \text{(A.23)}$$

[31] See Lemmas 3-6, pages 225-228.

and:

$$Z_n \equiv \sum_{k=0}^{n-1} \gamma_{k+1} \phi' (\theta_k) x_k \overline{Z}_k \tag{A.24}$$

Equipped with these definitions, we can make four crucial observations: (i) Z_n is a martingale with respect to the σ-algebra F_n generated by θ_0, Y_0, Y_1, ..., Y_n: u is a zero mean iid shock, which implies that $\overline{Z}_k$ is a martingale difference with respect to F_k; (ii) the following inequality holds:

$$I (n \leq \tau) \left| \sum_{k=0}^{n-1} \epsilon_k^1 (\phi) \right| \leq |Z_n| \tag{A.25}$$

and (iii) the fact that $\overline{Z}_k$ is a martingale difference with respect to F_k implies that[32]:

$$E |Z_n|^2 = \sum_{k=0}^{n-1} \gamma_{k+1}^2 E \phi' (\theta_k)^2 x_k^2 \overline{Z}_k^2 \tag{A.26}$$

Finally, (iv) we note that:

$$E x_k^2 \overline{Z}_k^2 \leq E \overline{x} (u_k)^2 \overline{Z}_k^2 \leq \widetilde{A}_1 (1 + |y|^{q_1}) \tag{A.27}$$

where we used the upper bound on the absolute value of the output gap in an optimum derived in the construction of the recursive representation of the CB problem, the assumption that u is an iid with finite moments and the fact that we are considering θ's inside a compact set.

We can combine these four observations with the Doob's martingale inequality as in Benveniste, Métivier, and Priouret (1990), Lemma 2 page 224, to conclude that:

$$E \left\{ \sup_{n \leq m} I (n \leq \tau) \left| \sum_{k=0}^{n-1} \epsilon_k^1 (\phi) \right| \right\}^2 \leq E \left\{ \sup_{n \leq m} |Z_n|^2 \right\} \leq 4 \sup_{n \leq m} E |Z_n|^2 \tag{A.28}$$

$$\leq A_1 (1 + |y|^{q_1}) \sum_{k=0}^{m-1} \gamma_{k+1}^2 \tag{A.29}$$

[32] See Evans and Honkapohja (1998), page 81 for the details.

hence proving the first part of the Lemma; note that we again used the fact that $\phi'(\theta)$ is a continuous function defined on a compact set, and hence has a maximum. The second part of the Lemma is a simple implication of the first one, and of the results derived to obtain it.[33] □

Lemma 5. *There exist constants A_4 and q_4 such that:*

$$E_{y,a}\left\{\sup_{n\leq m} I(n\leq\tau)\left|\sum_{k=0}^{n-1}\epsilon_k^4(\phi)\right|\right\}^2 \leq A_4\left(1+|y|^{q_4}\right)\sum_{k=0}^{m-1}\gamma_{k+1}^2 \tag{A.30}$$

Moreover, on $\{\tau\leq\infty\}$, $\sum_{k=0}^{n-1}\epsilon_k^1$ converges a.s. and in L^2.

Proof. The proof is analogous to the one of Lemma 4, once we note that:

$$I(k+1\leq\tau)\,\theta_k\Phi_1^2(b_k^\pi)\left(\sigma_u^2-u_{k+1}^2\right) \tag{A.31}$$

is a martingale difference with respect to F_k. □

Lemma 6. *There exist constants A_5 and q_5 such that:*

$$E_{y,a}\left\{\sup_{n\leq m} I(n\leq\tau)\left|\sum_{k=0}^{n-1}\epsilon_k^5(\phi)\right|\right\}^2 \leq A_5\left(1+|y|^{q_5}\right)\left(\sum_{k=0}^{m-1}\gamma_{k+1}^2\right)^2 \tag{A.32}$$

Proof. First of all, let's define:

$$D_n \equiv \sum_{k=0}^{n-1} I(k+1\leq\tau)\,\epsilon_k^5(\phi) \tag{A.33}$$

and note that:

$$\begin{aligned}\sup_{n\leq m} I(n\leq\tau)\left|\sum_{k=0}^{n-1}\epsilon_k^5(\phi)\right| \leq \sup_{n\leq m}|D_n| &\leq \sup_{n\leq m}\sum_{k=0}^{n-1} I(k+1\leq\tau)\left|\epsilon_k^5(\phi)\right| \\ &\leq \sum_{k=0}^{m-1} I(k+1\leq\tau)\left|\epsilon_k^5(\phi)\right| \end{aligned} \tag{A.34}$$

[33] See Benveniste, Métivier, and Priouret (1990), Lemma 2, page 225.

Moreover, we can use the same arguments used to derive polynomial bounds for the function $|\rho(\theta, Y)|$ to get:

$$\begin{aligned}
\left|\epsilon_k^5(\phi)\right| &= \left|\gamma_{k+1}\phi'(\theta_k)\,\theta_k\left(\Phi_2^2(s_{k+1}) + 2\Phi_2(s_{k+1})\,\Phi_1(b_k^\pi)\,u_{k+1}\right)\right| \\
&\leq \left|\gamma_{k+1}\phi'(\theta_k)\,\theta_k\left(\gamma_{k+1}^2\left(\frac{\Phi_2(s_{k+1})}{\gamma_{k+1}}\right)^2 + 2\gamma_{k+1}\frac{\Phi_2(s_{k+1})}{\gamma_{k+1}}\Phi_1(b_k^\pi)\,u_{k+1}\right)\right| \\
&\leq \left|\gamma_{k+1}^2\phi'(\theta_k)\,\theta_k\left(\left(\frac{\Phi_2(s_{k+1})}{\gamma_{k+1}}\right)^2 + 2\frac{\Phi_2(s_{k+1})}{\gamma_{k+1}}\Phi_1(b_k^\pi)\,u_{k+1}\right)\right| \\
&\leq \left|\gamma_{k+1}^2\phi'(\theta_k)\,\theta_k\widetilde{A}_5\left(1 + |Y_{k+1}|^{\overline{q}_5}\right)\right|
\end{aligned}$$

Putting these results together, and using the Cauchy-Schwarz inequality, we get:

$$\begin{aligned}
& E\left\{\sup_{n\leq m} I(n \leq \tau)\left|\sum_{k=0}^{n-1}\epsilon_k^5(\phi)\right|\right\}^2 \leq E\left\{\sum_{k=0}^{m-1} I(k+1 \leq \tau)\left|\epsilon_k^5(\phi)\right|\right\}^2 \\
\leq\ & \left(\sum_{k=0}^{m-1}\gamma_{k+1}^2\right)\left(\sum_{k=0}^{m-1}\gamma_{k+1}^2 E\left\{I(k+1 \leq \tau)\left|\phi'(\theta_k)\,\theta_k\widetilde{A}_5\left(1 + |Y_{k+1}|^{\overline{q}_5}\right)\right|^2\right\}\right) \\
\leq\ & A_5\left(1 + |y|^{q_5}\right)\left(\sum_{k=0}^{m-1}\gamma_{k+1}^2\right)^2
\end{aligned}$$

□

We can now state and prove our main result.

Proposition 5. *There exist constants A and q such that:*

$$E_{y,a}\left\{\sup_{n\leq m} I(n \leq \tau)\left|\sum_{k=0}^{n-1}\epsilon_k(\phi)\right|\right\}^2 \leq A\left(1 + |y|^q\right)\left(1 + \sum_{k=0}^{m-1}\gamma_{k+1}^2\right)\sum_{k=0}^{m-1}\gamma_{k+1}^2 \tag{A.35}$$

Moreover, on $\{\tau \leq \infty\}$, $\sum_{k=0}^{n-1}\epsilon_k^1$ converges a.s. and in L^2.

Proof. The decomposition of the error term $\epsilon(\phi)$ derived above, together with Lemmas 4-6 and the arguments in Benveniste, Métivier, and Priouret (1990), Lemmas 3-6, pages 225-228, imply that the first term in the inequality (A.35) is

bounded above by expressions of the form:

$$A_i \left(1 + |y|^{q_i}\right) \sum_{k=0}^{m-1} \gamma_{k+1}^2 \quad \text{or} \quad A_i \left(1 + |y|^{q_i}\right) \left(\sum_{k=0}^{m-1} \gamma_{k+1}^2 \right)^2 \tag{A.36}$$

By the Cauchy-Schwarz inequality, we have that:

$$\left(\sum_{k=0}^{m-1} \gamma_{k+1}^2 \right)^2 \leq \left(\sum_{k=0}^{m-1} \gamma_{k+1}^2 \right) \left(\sum_{k=0}^{m-1} \gamma_{k+1}^2 \right) \tag{A.37}$$

which implies that the inequality (A.35) holds. The second part of the Proposition is a trivial consequence of these upper bounds. □

Proof of Proposition 3. In the above Proposition we have established upper bounds on the fluctuations of the error term $\epsilon(\phi)$; in particular, our result is the exact counterpart of Proposition 7 of Benveniste, Métivier, and Priouret (1990), pages 228-229. The rest of the arguments leading to their convergence result (Theorem 13, page 236) go through also in our setup, so that we can conclude saying that *E-stability does imply (local) stability under learning in our model.* □