The suboptimality of commitment equilibrium when agents are learning *

(preliminary draft)

Antonio Mele[†], Krisztina Molnár[‡] and Sergio Santoro[§]

June 18, 2012

Abstract

Empirical research emphasizes that the impact of monetary policy on inflation expectations of private agents is sluggish. We explore the consequences of sluggish reaction of private expectations on the design of optimal monetary policy. We adopt a setup where the optimal monetary policy under commitment is welfare superior to the one under discretion if agents have rational expectations; moreover, if agents' beliefs slightly deviate from rational expectations, the central bank can still drive them to learn the rational expectations commitment equilibrium. We show that when private agents are learning a benevolent rational central bank does not implement the allocations consistent with rational expectation and commitment, not even asymptotically. The best policy is to make people learn the discretionary equilibrium instead.

JEL classification: C62, D83, D84, E52

Keywords: optimal monetary policy, learning, rational expectations, time consistency

---

*The views expressed herein are those of the authors, and do not necessarily reflect those of the Bank of Italy. We thank Andrea Caggese, Marco Del Negro, John Duca, Tore Ellingsen, Martin Ellison, Michal Horvath, Albert Marcet, Ramon Marimon, Andrzej Nowak, Aarti Singh for useful comments.

[†]University of Oxford; Email: antonio.mele@economics.ox.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

1 Introduction

Since Kydland and Prescott (1977) and Barro and Gordon (1983) it is widely accepted that managing private expectations is one of the most important tasks of central banks. This have had an enormous impact on the practice of central banking. Yet, there still seems to be a divide between theory and practice. Most theoretical models suggest that monetary policy has an immediate impact on private expectations, while empirical research, analyzing actual central bank practice suggests that this impact is sluggish.[1]

This paper explores the consequences of sluggish private expectations on the design of optimal monetary policy. We use a standard monetary model which features a trade-off between inflation and output gap stabilization. It is well known that when the private sector has rational expectations (RE), a credible central bank can decrease the inflation expectations of agents by committing to a series of economic contractions; the equilibrium induced by commitment and RE (RECE) leads to significant welfare gains compared to the RE discretionary equilibrium (REDE).[2] Our main question is to examine whether the the allocations of the RECE remain optimal even if private expectations slightly deviate from rationality.

We assume agents' expectations deviate from full rationality but we keep the assumption that the bank is able to decrease inflation expectations by engineering output contractions. The main difference with respect to rational agents is that our agents are continuously learning about how output contractions affect inflation. Therefore the central bank can "train" agents to understand how economic contractions lead to lower inflation by engineering a series of equilibrium allocations consistent with the RECE. In other words, when agents are rational, the central bank can impact expectations through promises, while in our framework the central bank can only impact expectations by consistency of actions.

Evans and Honkapohja (2006) and Evans and Honkapohja (2003) have shown that when private agents are learning about equilibrium allocations, both RECE and REDE are attainable. Our paper extends their research by posing a normative question, and asking which equilibrium is desirable under learning. We assume private agents can learn the allocations of both RECE and REDE, it is up to the benevolent central bank to decide which equilibrium to choose.

We show that the welfare ordering of equilibria is different under RE and learning. Under learning, optimal monetary policy drives the economy far from the RE commitment equilibrium, and to the RE discretionary equilibrium. This holds true even if initial beliefs are consistent with RECE.

The intuition behind this surprising result comes from the sluggish nature

[1] See for example Haldane and Read (2000) and Thornton (2003).
[2] See Gali (2003), Woodford (2003).

of inflation expectations. Substituting RE with learning does not influence the stabilization bias: in the long run, being in the RECE the central bank is able to counteract an adverse supply shock with a smaller welfare loss than being in the REDE. However in the short run the central bank can gain a substantial welfare increase by exploiting the sluggish nature of private expectations. Compared to optimal policies derived assuming RE, optimal policy under learning engineers much bigger output contractions, and this in turn will generate a bigger decrease in inflation expectations and inflation. This short term gain of being able to decrease inflation more gives such a big welfare gain, that it compensates for the long run losses. This mechanism is similar to the old literature on adaptive expectations, but an important difference is that learning agents cannot be fooled forever. As the central banks deviates from the optimal RECE allocations, so will private agents' beliefs. In the long run expectations converge to the REDE allocations and the central bank looses its ability to impact inflation expectations through output contractions.

We follow the methodology of Gaspar, Smets, and Vestin (2006) and Molnar and Santoro (2010), and determine optimal policy assuming the central bank knows and makes active use of the exact form of private expectations. The central bank, in other words, takes into account the effect of its policy choices on the expectation formation process, and exploits it.

Note, that since agents form beliefs by running regressions on past data, the central bank does not have a time inconsistency problem: current (and past) policy determines future expectations. In other words, when the central bank internalizes the learning process of the agents, the optimal allocation is a Markov perfect equilibrium. Such an equilibrium cannot be consistent with asymptotic convergence to RECE, since the latter is not a Markov equilibrium. The optimal path is Markov perfect, and it converges to the unique Markov perfect RE equilibrium in this economy, the REDE.

The early literature often motivated learning of private agents as a device for selecting among multiple equilibria: if agents' expectations are perturbed out of an equilibrium, are they able to converge back, or learn another one? In these papers equilibrium stability depends on the stability of the learning algorithm, the so called E-stability (see Marcet and Sargent (1989), Evans and Honkapohja (2001)). Our paper extends this literature by examining equilibrium selection when learners interact with an optimizing rational agent. The resulting equilibrium does not only depend on the stability properties of the learners, but also on the incentives of the rational agents.

A subtle, but important difference between learning agents and rational agents is that the latter understand off-equilibrium strategies, while learning agents cannot. As Chari and Kehoe (1990) and Kurozumi (2008) show, if the central bank

(CB) is patient enough the RECE is a sustainable equilibrium even if there is no commitment device, but the CB takes into account certain incentive compatibility constraints that constitute an off-equilibrium threat of private agents. Would the bank deviate, so would private agents, and this makes deviation suboptimal. In our setup, contrary to Chari and Kehoe (1990) and Kurozumi (2008) the game is not between two rational players, but a rational and a learning agent. The learning agent cannot posit credible threats, and can only change their actual behavior sluggishly. As a result the bank remains with a strong incentive to deviate from RECE, because of its time inconsistent nature.

Our research is most closely related to Sargent (1999), chapter 5, in which a rational CB exploits the (mechanical) forecasting rule of the agents in a model with natural rate of unemployment (the so called Phelps problem). Sargent (1999) shows that a patient enough CB can asymptotically replicate the commitment solution under RE. Also Molnar and Santoro (2010) and Gaspar, Smets, and Vestin (2006) show that there are some qualitative similarities between the learning solution and the commitment solution under RE. However, they cannot make a formal argument like Sargent (1999). The reason for this is that in the Phelps problem considered in Sargent (1999) the equilibrium under discretion and commitment have the same functional form, while in Molnar and Santoro (2010) and Gaspar, Smets, and Vestin (2006) this is not true. In Sargent (1999) the discretionary and the commitment solution under RE are a constant. Hence, agents are learning the value of a constant, and they have the "possibility" to learn either discretion or commitment equilibria. In the New Keynesian model, used by Molnar and Santoro (2010) and Gaspar, Smets, and Vestin (2006), discretion and commitment solution under RE have different functional forms. Since they assume that the agents' beliefs about policy have the functional form of the RE solution under discretion, they cannot replicate the experiment carried out by Sargent: the commitment solution has a very different form. In this paper we make a step further, and posit that agents are learning in a form that is consistent with both RECE and REDE. Therefore, by choosing the optimal policy, the central bank can drive the economy either in one or the other.

One implication of our result is that RECE might not be optimal in the long run. Other researchers questioned the importance of the commitment equilibrium on different grounds. Levin, Wieland, and Williams (1999) show that commitment policies can perform very poorly if the central bank's reference model is badly misspecified. Orphanides and Williams (2008) shows similar results when the central bank has misspecified belief about private expectations.

2 The Model

We consider the baseline version of the New Keynesian model; in this framework, the economy is characterized by two structural equations.[3] 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 the private agents' conditional expectation, which is not necessarily rational. The above equation is derived by loglinearizing 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[4]; this relation is obtained from optimal pricing decisions of monopolistically competitive firms whose prices are staggered à la Calvo (1983).[5]

The loss function of the CB is given by:

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

where α is the relative weight put by the CB on the objective of output gap stabilization.[6]

---

[3] 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).

[4] 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.

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

[6] As is shown in Rotemberg and Woodford (1997), equation (3) can be obtained as a quadratic approximation to the expected household's utility function; in this case, α is a function of structural parameters.

2.1 Commitment and discretion solution under RE

Assume that the private sector has RE, 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:

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

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

$$\pi_0 = -\frac{\alpha}{\kappa} x_0 \tag{5}$$
$$\pi_t = -\frac{\alpha}{\kappa} x_t + \frac{\alpha}{\kappa} x_{t-1}, \quad t \geq 1 \tag{6}$$

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). To overcome this problem, Woodford (2003) proposed to adopt the optimal policy (5)-(6) from a "timeless perspective", namely from such a long distance from the moment in which the optimization is carried out that we can apply (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:

$$b^x = \frac{\kappa^2 + \alpha(1+\beta) - \sqrt{(\kappa^2 + \alpha(1+\beta))^2 - 4\alpha^2\beta}}{2\alpha\beta} \tag{9}$$
$$b^\pi = \frac{\alpha}{\kappa}\left(1 - b^x\right) \tag{10}$$
$$c^x = -\frac{\kappa b^x}{\alpha} \tag{11}$$
$$c^\pi = -\frac{\alpha}{\kappa} c^x \tag{12}$$

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 characterized by

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

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

2.2 Learning specification

In the rest of the paper, we dispose of the assumption that the private sector has RE. Following Molnar and Santoro (2010), we posit that the central bank is fully rational. However, we assume that private agents are adaptive learners. In particular, they know the structure of the economy, but they do not know the parameters' values. Hence, they estimate them by observing past and current allocations.[7]

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 the central bank would implement under RE and commitment (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 least squares learning, agents estimate equations (16)-(17) and use the estimates $\left(b^\pi_{t-1}, b^x_{t-1}\right)$ to make forecasts:

$$E^*_t \pi_{t+1} = b^\pi_{t-1} x_t \tag{18}$$

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

We can interpret this assumption as agents understanding that the central bank is committed, but not knowing the exact quantitative impact of central bank's actions on equilibrium allocations.

[7] The modern literature on adaptive learning was initiated by Marcet and Sargent (1989), 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).

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:

$$E_t^* \pi_{t+1} = b_{t-1}^{\pi} \left(b_{t-1}^{x} x_{t-1} + c_{t-1}^{x} u_t \right) \quad (20)$$

$$E_t^* x_{t+1} = b_{t-1}^{x} \left(b_{t-1}^{x} x_{t-1} + c_{t-1}^{x} u_t \right). \quad (21)$$

Since expectations depend also on the estimated values of the coefficients c^{π} and c^{x}, an optimizing CB should take into account their updating algorithms as well. This way we would end up having two more state variables, with significant additional complications in the numerical exercise.

We assume the coefficients are estimated with stochastic gradient learning (this basically means that we abstract from the evolution of the estimated second moments of the regressors). The recursive formulation of the regression coefficients is the following:

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

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

We focus on two different learning algorithms. The first is decreasing gain learning, where the gain parameter $\gamma_t = \frac{1}{t}$. In fact, this algorithm corresponds to the case in which agents run least squares regressions on past data to estimate the parameters of the PLM. The second algorithm is constant gain learning, where $\gamma_t = \gamma \in (0, 1)$.

3 An heuristic presentation of the result

The problem of the central bank is to minimize the welfare loss function (3) subject to the IS curve (1), the New Keynesian Phillips curve (2), and the learning updating process (23)-(22). In the next section, we formally prove our main result: under decreasing gain learning, the optimal policy drives the economy towards REDE. However, in this section we present a non-technical, simplified analysis that highlights the main effects at work. In particular, we disentangle three effects: an intratemporal stabilization trade-off, an intertemporal smoothing effect and an intertemporal learning cost. In the next section, we show that the first order conditions for an optimum are

$$
\begin{aligned}
0 = & -\alpha x_t - \left[(\beta b^{\pi}_{t-1} + \kappa)x_t + u_t\right](\beta b^{\pi}_{t-1} + \kappa) - \lambda_{1,t}\gamma x_{t-1}(\beta b^{\pi}_{t-1} + \kappa) - \qquad (24)\\
& - E_t[\lambda_{1,t+1}\beta\gamma((\beta b^{\pi}_t + \kappa)x_{t+1} + u_{t+1} - b^{\pi}_t 2x_t)] \qquad (25)\\
0 = & \lambda_{1,t} - \beta E_t \lambda_{1,t+1}(1 - \gamma x_t^2) - \beta^2 E_t[((\beta b^{\pi}_t + \kappa)x_{t+1} + u_{t+1})\, x_{t+1}] - \\
& \beta^2 E_t\left[\lambda_{1,t+1}\gamma x_t x_{t+1}\right] \qquad (26)
\end{aligned}
$$

To gain intuition lets assume that learning is shut down: $\gamma = 0$ and agents have initial beliefs different from the RECE coefficient $b^{\pi}_{t-1} \neq b^{\pi}$. As a result of $\gamma = 0$ agents keep their belief at b_{t-1} forever. The equations (25)-(26) yield:

$$x_t = -\frac{\beta b^{\pi}_{t-1} + \kappa}{\alpha + \left(\beta b^{\pi}_{t-1} + \kappa\right)^2} u_t \qquad (27)$$

Combining this with the Phillips curve, we obtain that

$$\pi_t = \frac{\alpha}{\alpha + \left(\beta b^{\pi}_{t-1} + \kappa\right)^2} u_t \qquad (28)$$

Comparing (28) to the inflation allocation under REDE (i.e. equation (15)) we can see that the optimal policy under learning is very similar to the discretionary policy. In fact, by setting $b^{\pi}_{t-1} = 0$ we are back to the same policy under REDE, and this isolates the classic *intratemporal stabilization trade-off*: the presence of a cost push shock does not allow the central bank to set both zero inflation and zero output gap. Hence when setting the optimal policy, the central bank must follow a "leaning against the wind" strategy: after a positive shock, it decreases current output gap in order to avoid a huge increase in current inflation.

Assume now $b^{\pi}_{t-1} > 0$ (the negative case is similar). There is a second effect that the monetary authority must take into account, which we call the *intertemporal smoothing effect*. The interpretation is straightforward: when setting the optimal policy, the central bank still follows a "leaning against the wind" strategy as in REDE. However, even if beliefs are fixed at b^{π}_{t-1}, inflation expectations are not given: the central bank can influence them by choosing the output gap, since $E_t\pi_{t+1} = b^{\pi}_{t-1}x_t$. Therefore, the benevolent monetary authority has to take into account that the output gap today will determine the expected inflation tomorrow. A lower expected inflation tomorrow allows the central bank to keep inflation low today, therefore smoothing the effect of the shock between today's and tomorrow's inflation. In order to do that, the reaction of output gap to the cost push shock must be larger than under REDE. In particular, the larger is b^{π}_{t-1}, the stronger is effect of a decrease in output gap on future inflation, and therefore the better the central bank can smooth stabilization intertemporally. When inflation expectations decrease in response to an output contraction, $b^{\pi}_{t-1} > 0$, the bank's ability to intertemporally smooth out cost push shocks is reflected in a decreased volatility of inflation: (28) yields lower volatility than (15).

Notice the similarity with RECE, where the central bank must commit to an infinite sequence of future choices for output gap in order to smooth the cost of a shock today. In RE, the NKPC can be solve forward to get

$$\begin{aligned} \pi_t &= \beta E_t \pi_{t+1} + \kappa x_t + u_t \\ &= \kappa \sum_{k=1}^{\infty} \beta^k E_t \left[x_{t+k} \right] + \kappa x_t + u_t \end{aligned}$$

Therefore, when a positive shock hits, the optimal policy under RECE is to slightly rise inflation in the current period while decreasing current output gap and committing to reduce output gap in the future. This commitment has a cost, which is summarized in the past shadow value of the NKPC constraint, i.e. the Lagrange multiplier associated with it. The benefit, however, is a better intertemporal allocation of the fluctuations, and therefore higher welfare.

A difference to RECE is that in our framework without learning, this cost is absent, since the optimal policy does not involve committing to future output gap sequences: inflation expectations are anchored by the choice of current output gap and the current beliefs (which are given).

In sum, *intertemporal smoothing effect* for $b_{t-1} > 0$ and $\gamma = 0$ follows from the fact that inflation expectations react to output contractions in a similar fashion as in the RECE and as a result optimal policy is able to smooth out intertemporally the effect of cost push shocks in a similar fashion as optimal policy in the RECE.

Finally, there is a another intertemporal effect that is related to the learning process. This is a cost that does not arise in the RECE, only under learning, therefore we call this the *intertemporal learning cost*. Assume now learning is back at work, i.e. $\gamma \neq 0$. For the sake of intuition let $b^{\pi}_{t-1} > 0$. Then agents will respond to optimal policy by reducing the absolute value of their learning coefficients. This constitutes a cost: a lower b^{π}_{t-1} implies that is more difficult to exploit the smoothing trade-off. To show this analytically, we can calculate the expected value of the learning coefficients. For a small γ , by a continuity argument we can use the policy function (27)-(28), and approximate the learning algorithm as

$$\begin{aligned} b^{\pi}_t =& b^{\pi}_{t-1} + \gamma x_{t-1} \left(\pi_t - x_{t-1} b^{\pi}_{t-1} \right) \\ \simeq& b^{\pi}_{t-1} + \gamma x_{t-1} \left(\frac{\alpha}{\alpha + \left(\beta b^{\pi}_{t-1} + \kappa \right)^2} u_t - x_{t-1} b^{\pi}_{t-1} \right) \\ \simeq& b^{\pi}_{t-1} + \gamma x_{t-1} \frac{\alpha}{\alpha + \left(\beta b^{\pi}_{t-1} + \kappa \right)^2} u_t - \gamma x^2_{t-1} b^{\pi}_{t-1}. \end{aligned}$$

Taking expectations of b_t at time t we get:

$$\begin{aligned} E_{t-1}\left[b_t^\pi\right] \simeq & b_{t-1}^\pi + \gamma x_{t-1} \frac{\alpha}{\alpha + \left(\beta b_{t-1}^\pi + \kappa\right)^2} E_t\left(u_t\right) - \gamma x_{t-1}^2 b_{t-1}^\pi \\ \simeq & b_{t-1}^\pi\left(1 - \gamma x_{t-1}^2\right), \end{aligned} \tag{29}$$

hence $E_{t-1}\left[b_t^\pi\right] < b_{t-1}^\pi$, i.e. b_t behaves almost like a supermartingale when γ is very small, and hence tends to get closer to zero.[8] In other words, the learning algorithm on average reduces b_{t-1}^π and therefore increases the variance of inflation. If we calculate the instantaneous welfare loss we obtain

$$\begin{aligned} E_{t-1}\left[\pi_t^2 + \alpha x_t^2\right] &= E_{t-1}\left[\left(\frac{\alpha}{\alpha + \left(\beta b_{t-1}^\pi + \kappa\right)^2} u_t\right)^2 + \alpha\left(-\frac{\beta b_{t-1}^\pi + \kappa}{\alpha + \left(\beta b_{t-1}^\pi + \kappa\right)^2} u_t\right)^2\right] \\ &= \frac{\alpha}{\alpha + \left(\beta b_{t-1}^\pi + \kappa\right)^2} \sigma_u^2 \end{aligned}$$

The *cost of learning* is thus that it worsens the bank's ability to smooth out the effect of shocks across time : by having a lower b_{t-1}^π, the same output gap jump produces a smaller effect on expected inflation and therefore the stabilization has to rely more on increases in current inflation. This results in higher current welfare losses. However, short term gains are larger.

3.1 Evans and Honkapohja (2006)'s policy

We can easily see why the optimal policy under learning is Pareto superior to the one suggested in Evans and Honkapohja (2006), which drives the economy towards RECE:

$$x_t = \frac{\alpha}{\alpha + \kappa\left(\beta b_{t-1}^\pi + \kappa\right)} x_{t-1} - \frac{\kappa}{\alpha + \kappa\left(\beta b_{t-1}^\pi + \kappa\right)} u_t$$

and

$$\pi_t = \frac{\alpha\left(\beta b_{t-1}^\pi + \kappa\right)}{\alpha + \kappa\left(\beta b_{t-1}^\pi + \kappa\right)} x_{t-1} + \frac{\alpha}{\alpha + \kappa\left(\beta b_{t-1}^\pi + \kappa\right)} u_t.$$

[8] The case with $b_{t-1}^\pi < 0$ is symmetric and equation (29) implies that the coefficient becomes less negative and closer to zero.

Assume again that learning is shut down. Hence, we can calculate the expected instantaneous welfare loss as

$$\begin{aligned}
E_{t-1}\left[\pi_t^2+\alpha x_t^2\right] &= E_{t-1}\left[\left(\frac{\alpha\left(\beta b_{t-1}^{\pi}+\kappa\right)}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}x_{t-1}+\frac{\alpha}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}u_t\right)^2+\right. \\
&\quad \left.\alpha\left(\frac{\alpha}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}x_{t-1}-\frac{\kappa}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}u_t\right)^2\right] \\
&= \left(\frac{\alpha\left(\beta b_{t-1}^{\pi}+\kappa\right)}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}\right)^2 x_{t-1}^2+\left(\frac{\alpha}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}\right)^2\sigma_u^2+ \\
&\quad \alpha\left[\left(\frac{\alpha}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}\right)^2 x_{t-1}^2+\left(\frac{\kappa}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}\right)^2\sigma_u^2\right] \\
&= \frac{\alpha^2\left[\left(\beta b_{t-1}^{\pi}+\kappa\right)^2+\alpha\right]}{\left[\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)\right]^2}x_{t-1}^2+\frac{\alpha\left[\alpha+\kappa^2\right]}{\left[\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)\right]^2}\sigma_u^2
\end{aligned}$$

Let us compare it with the optimal policy. Notice that EH policy has an extra term that depends on the variability of output gap. Therefore, if the variability directly induced by the cost push shock would be the same, the EH policy would be more costly. We therefore look at the variability induced by cost push shocks only. Let us denote the ratio of the coefficients as R

$$R\equiv\frac{\frac{\alpha\left[\alpha+\kappa^2\right]}{\left[\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)\right]^2}}{\frac{\alpha}{\alpha+\left(\beta b_{t-1}^{\pi}+\kappa\right)^2}}=\left(\alpha+\kappa^2\right)\frac{\left(\alpha+\left(\beta b_{t-1}^{\pi}+\kappa\right)^2\right)}{\left[\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)\right]^2}$$

We can do some simple analysis. If $b_{t-1}^{\pi}=0$, then we have

$$R=\left(\alpha+\kappa^2\right)\frac{\left(\alpha+\kappa^2\right)}{\left[\alpha+\kappa^2\right]^2}=1$$

i.e. the two coefficients are the same in REDE. Therefore, when the economy approaches the discretionary equilibrium, EH policy induces larger variability because of its dependence on output gap variance. What happens if the economy is close to the commitment equilibrium? The derivative with respect to b_{t-1}^{π} is

$$\frac{\partial R}{\partial b_{t-1}^{\pi}}\equiv\frac{2\left(\alpha+\kappa^2\right)\alpha\beta^2 b_{t-1}^{\pi}}{\left[\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)\right]^3}$$

which is always positive as long as b^{π}_{t-1} is positive. Since the derivative of R is positive, then the variability induced by the cost push shock will be higher for EH policy. In fact we have

$$R = \left(\alpha + \kappa^2\right) \frac{\left(\alpha + \left(\beta b^{\pi}_{COM} + \kappa\right)^2\right)}{\left[\alpha + \kappa \left(\beta b^{\pi}_{COM} + \kappa\right)\right]^2}$$

Figure 1 shows that the ratio R is always larger than 1 for reasonable values of the parameters. Therefore, the coefficient of the EH policy is always larger and the EH policy induces higher volatility. Now let learning be back in the picture.

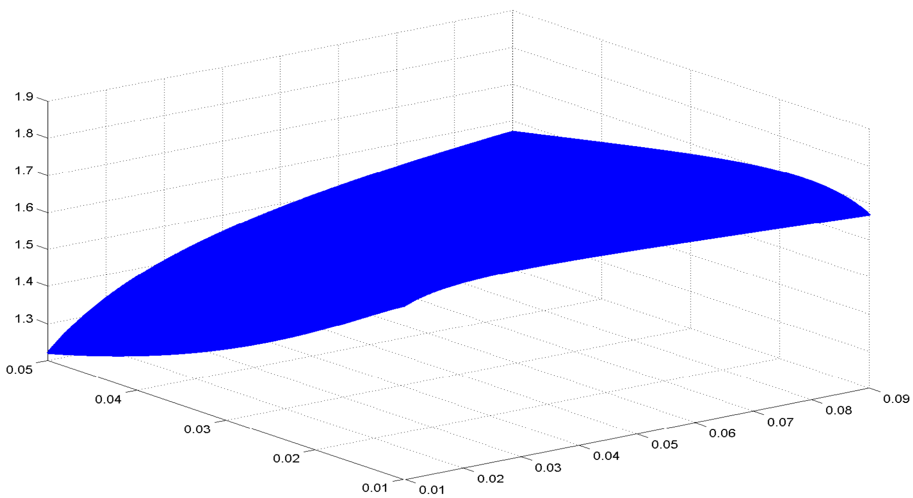

Figure 1: The ratio R under different combinations of κ and α

Learning however implies

$$\begin{aligned} b^{\pi}_t =& b^{\pi}_{t-1} + \gamma x_{t-1}\left(\pi_t - x_{t-1} b^{\pi}_{t-1}\right) \\ =& b^{\pi}_{t-1} + \gamma x_{t-1}\left(\frac{\alpha\left(\beta b^{\pi}_{t-1} + \kappa\right)}{\alpha + \kappa\left(\beta b^{\pi}_{t-1} + \kappa\right)} x_{t-1} + \frac{\alpha}{\alpha + \kappa\left(\beta b^{\pi}_{t-1} + \kappa\right)} u_t - x_{t-1} b^{\pi}_{t-1}\right) \\ =& b^{\pi}_{t-1} + \frac{\alpha\left(\beta b^{\pi}_{t-1} + \kappa\right)}{\alpha + \kappa\left(\beta b^{\pi}_{t-1} + \kappa\right)} \gamma x^2_{t-1} + \gamma x_{t-1} \frac{\alpha}{\alpha + \kappa\left(\beta b^{\pi}_{t-1} + \kappa\right)} u_t - \gamma x^2_{t-1} b^{\pi}_{t-1} \end{aligned}$$

and taking expectations

$$E_{t-1}\left[b^{\pi}_t\right] = b^{\pi}_{t-1} + \gamma x^2_{t-1}\left(\frac{\alpha\left(\beta b^{\pi}_{t-1} + \kappa\right)}{\alpha + \kappa\left(\beta b^{\pi}_{t-1} + \kappa\right)} - b^{\pi}_{t-1}\right)$$

which implies that the expected coefficient is higher if

$$\frac{\alpha\left(\beta b_{t-1}^{\pi}+\kappa\right)}{\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)}-b_{t-1}^{\pi}>0$$

or

$$b_{t-1}^{\pi}\left(\alpha+\kappa\left(\beta b_{t-1}^{\pi}+\kappa\right)\right)-\alpha\left(\beta b_{t-1}^{\pi}+\kappa\right)<0$$

Rearranging:

$$\kappa\beta\left(b_{t-1}^{\pi}\right)^{2}+\left(\alpha\left(1-\beta\right)+\kappa^{2}\right)b_{t-1}^{\pi}-\alpha\kappa<0$$

The roots are

$$b=\frac{-\left(\alpha\left(1-\beta\right)+\kappa^{2}\right)\pm\sqrt{\left(\alpha\left(1-\beta\right)+\kappa^{2}\right)^{2}+4\alpha\beta\kappa^{2}}}{2\kappa\beta}$$

or

$$b=\frac{\alpha}{\kappa}\left(1-\frac{\kappa^{2}+\alpha(1+\beta)\mp\sqrt{(\kappa^{2}+\alpha(1+\beta))^{2}-4\alpha^{2}\beta}}{2\alpha\beta}\right)$$

where the positive root is the coefficient for RECE. Therefore, the expected coefficient will be higher if

$$\begin{aligned}b_{t-1}^{\pi}\ \in\ & \left[\frac{\alpha}{\kappa}\left(1-\frac{\kappa^{2}+\alpha(1+\beta)+\sqrt{(\kappa^{2}+\alpha(1+\beta))^{2}-4\alpha^{2}\beta}}{2\alpha\beta}\right),\right.\\ & \left.\frac{\alpha}{\kappa}\left(1-\frac{\kappa^{2}+\alpha(1+\beta)-\sqrt{(\kappa^{2}+\alpha(1+\beta))^{2}-4\alpha^{2}\beta}}{2\alpha\beta}\right)\right]\end{aligned}$$

In other words, on average the learning process increases the volatility more than under the optimal policy when we approach RECE. The EH policy brings the economy towards the RECE. In doing so, it overstates the benefits of the long run stabilization versus the short term gains, and incurs in additional costs coming from the learning process. Therefore, it increases the volatility of the economy and decreases welfare.

4 Decreasing gain learning

In this section, we study the economy under decreasing gain learning. 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 the main convergence result: under the optimal policy, the REDE is stable under learning.

4.1 Recursivity

We start stating the control problem of the CB 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.

$$\begin{aligned}
&\sup_{\{\pi_t, x_t, r_t, b_t^\pi, b_t^x\}_{t=0}^\infty} E_0(1-\beta)\sum_{t=0}^\infty \beta^t \left[-\frac{1}{2}\left(\pi_t^2 + \alpha x_t^2\right)\right] \\
&\text{s.t.} \\
&x_t = \frac{-\sigma^{-1} r_t}{1 - b_{t-1}^x - \sigma^{-1} b_{t-1}^\pi} \\
&\pi_t = (\beta b_{t-1}^\pi + \kappa) x_t + u_t \\
&b_t^\pi = b_{t-1}^\pi + \gamma_t x_{t-1}\left(\pi_t - x_{t-1} b_{t-1}^\pi\right) \\
&b_t^x = b_{t-1}^x + \gamma_t x_{t-1}\left(x_t - x_{t-1} b_{t-1}^x\right), \\
&x_{-1}, b_{-1}^\pi, b_{-1}^x, \gamma_0 \text{ given}
\end{aligned}$$

Since the IS curve is never a binding constraint (the CB can always use to the interest rate to satisfy it), and using the NKPC to substitute out π, the above problem can be written in a simpler form:

$$\sup_{\{x_t, b_t^\pi, b_t^x\}_{t=0}^\infty} E_0(1-\beta)\sum_{t=0}^\infty \beta^t \left\{-\frac{1}{2}\left[\left((\beta b_{t-1}^\pi + \kappa)x_t + u_t\right)^2 + \alpha x_t^2\right]\right\} \tag{30}$$

s.t.

$$b_t^\pi = b_{t-1}^\pi + \gamma_t x_{t-1}\left((\beta b_{t-1}^\pi + \kappa)x_t + u_t - x_{t-1} b_{t-1}^\pi\right) \tag{31}$$

$$b_t^x = b_{t-1}^x + \gamma_t x_{t-1}\left(x_t - x_{t-1} b_{t-1}^x\right), \tag{32}$$

$$x_{-1}, b_{-1}^\pi, b_{-1}^x, \gamma_0 \text{ given} \tag{33}$$

In any period t the state variables in the above problem are five: three endogenous $(x_{t-1}, b_{t-1}^\pi, b_{t-1}^x)$ that take values in $\mathbb{R}^3$, one exogenous and stochastic (u_t) defined on some underlying probability space and that takes values in a measurable space $(Z, \mathfrak{Z})$, and one exogenous and deterministic (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 CB 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:

$$\text{for any } s \in S,\ \Gamma(s) = \left\{a \in \mathbb{R}^3 : \text{equations (31) and (32) hold}\right\}$$

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 (30). To do so, the strategy we adopt is the following: first of all, 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 made so big 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:

$$\sup_{\{x_t, b_t^\pi, b_t^x\}_{t=0}^\infty} E_0(1-\beta)\sum_{t=0}^{\infty}\beta^t\left\{-\frac{1}{2}\left[\left((\beta b_{t-1}^\pi+\kappa)x_t+u_t\right)^2+\alpha x_t^2\right]\right\} \tag{34}$$

s.t.

$$b_t^\pi = b_{t-1}^\pi + \gamma_t x_{t-1}\left((\beta b_{t-1}^\pi+\kappa)x_t+u_t-x_{t-1}b_{t-1}^\pi\right) \tag{35}$$

$$b_t^x = b_{t-1}^x + \gamma_t x_{t-1}\left(x_t - x_{t-1}b_{t-1}^x\right), \tag{36}$$

$$\overline{x}\left(s_t\right) \geq x_t \geq -\overline{x}\left(s_t\right), \tag{37}$$

$$x_{-1}, b_{-1}^\pi,\, b_{-1}^x, \gamma_0 \text{ given} \tag{38}$$

where the function of the states $\overline{x}\left(s_t\right)$ is an arbitrary continuous function. Let's now fix some notation. The vector of the state variables at time t is $s_t \equiv [x_{t-1}, b_{t-1}^\pi, b_{t-1}^x, u_t, \gamma_t]'$, while the vector of choice variables at t is $a_t \equiv [x_t, b_t^\pi, b_t^x]'$. 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:

$$\begin{array}{rcl} s_{t+1}^1 & = & a_t^1 \\ s_{t+1}^2 & = & a_t^2 \\ s_{t+1}^3 & = & a_t^3 \\ s_{t+1}^4 & = & \xi \\ s_{t+1}^5 & = & \dfrac{s_t^5}{1+s_t^5} \end{array}$$

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:

$$s_{t+1} = \Psi\left(s_t, a_t, \xi\right) \tag{39}$$

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:

$$Q\left(D|s,a\right) = \int_Z \mathbf{1}_D\left(\Psi\left(s,a,\xi\right)\right) dP\left(\xi\right) \tag{40}$$

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:*

$$\textit{for any } s \in S,\ \Gamma^c\left(s\right) = \left\{a \in \mathbb{R}^3 : \textit{equations } (35), (36) \textit{ and } (37) \textit{ hold}\right\}$$

is compact-valued.

(ii) *The feasibility correspondence:*

$$\textit{for any } s \in S,\ \Gamma^c\left(s\right) = \left\{a \in \mathbb{R}^3 : \textit{equations } (35), (36) \textit{ and } (37) \textit{ hold}\right\}$$

is upper hemi-continuous.

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

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

is continuous.

Proof. **(i)** For any value of $s \in S$, equation (35) is a linear function of b_t^π and x_t, and analogously equation (36) 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:

$$\Gamma^c(s) \subset [-\overline{x}(s), \overline{x}(s)] \times [\underline{b}^\pi(s), \overline{b}^\pi(s)] \times [\underline{b}^x(s), \overline{b}^x(s)] \tag{41}$$

Moreover, by linearity (conditional on s) of the equations (35) and (36), 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 (35) and (36) (they are formed by sums and products of elements of $\{s_n\}$ and $\{x_n\}$) implies that if the subsequences $\{b_{n_k}^\pi\}$ and $\{b_{n_k}^x\}$ satisfy equations (35) and (36) for any n_k, then they converge and the limit satisfies (35) and (36) 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 (40) 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 34.*

Proof. This result follows from Theorem 1 of Jaskiewicz and Nowak (2011). 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) 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 one strategy allows us to the apply their results to our model. □

Next, we prove that any optimal time-invariant policy function for the problem 34 is such that the constraint (37) never binds in the optimum, if an appropriate continuous function $\overline{x}(s)$ is chosen. We define $V^c(s)$ the value function associated to the solution of the problem 34 for a given initial vector of states $s \in S$.[9] In the following simple Lemma we characterize bounds of this value function.

Lemma 2. *Let's 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.

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

(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.

□

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 34 is such that the constraint* (37) *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:

$$V^c(s) = -(1-\beta)\frac{1}{2}\left[(\beta b^{\pi}+\kappa)x^*(s)+u)^2+\alpha x^{*2}(s)\right]+\beta\int_S V^c(s)\,Q\left(dy|s,a^*(s)\right) \tag{42}$$

for any $s \in S$, where the starred variables denotes actions taken under any optimal policy function. Using Lemma 2 (i) and the fact that $-(1-\beta)\frac{1}{2}(\beta b^{\pi}+\kappa)x^*(s)+u)^2$ 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 $x^*(s) = \overline{x}(s)$.[10] This means that:

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

which implies:

$$V^c(s) \leq -(1-\beta)\frac{1}{2}\alpha x^{*2}(s) < -\frac{1}{2}\left[(1-\beta)u^2+\beta\sigma_u^2\right] \tag{43}$$

which contradicts Lemma 2 (ii). □

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

4.2 Convergence

So far we proved that there exists an optimal time-invariant solution to the problem 34 and that it is interior; hence, any such solution can be characterized as the solution of the standard FOCs, without having to worry about the Lagrange multipliers on the constraints (37). The first order conditions of problem 34 are:

$$\begin{aligned} 0 =& -\alpha x_t - \left[(\beta b^\pi_{t-1}+\kappa)x_t + u_t\right](\beta b^\pi_{t-1}+\kappa) - \lambda_{1,t}\gamma_t x_{t-1}(\beta b^\pi_{t-1}+\kappa)- \qquad (44)\\ & - E_t[\lambda_{1,t+1}\beta\gamma_{t+1}((\beta b^\pi_t+\kappa)x_{t+1}+u_{t+1}-b^\pi_t 2x_t)] - \lambda_{2,t}\gamma_t x_{t-1}\\ & - E_t[\lambda_{2,t+1}\beta\gamma_{t+1}(x_{t+1}-b^x_t 2x_t)]\\ 0 =& \lambda_{1,t} - \beta E_t\lambda_{1,t+1}(1-\gamma_{t+1}x_t^2) - \beta^2 E_t[((\beta b^\pi_t+\kappa)x_{t+1}+u_{t+1})\,x_{t+1}]- \qquad (45)\\ & \beta^2 E_t\left[\lambda_{1,t+1}\gamma_{t+1}x_t x_{t+1}\right]\\ 0 =& \lambda_{2,t} - \beta E_t\lambda_{2,t+1}(1-\gamma_{t+1}x_t^2), \qquad (46) \end{aligned}$$

where $\lambda_{1,t}$ and $\lambda_{2,t}$ are the Lagrange multipliers of (35) and (36), respectively. These first order conditions together with the law of motion for the learning coefficients constitute the necessary conditions for the optimal evolution of $\{x_t, b^\pi_t, b^x_t\}$.[11] From equation (44) 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:

$$\begin{aligned} 0 =& -\alpha x_t - \left[(\beta b^\pi_{t-1}+\kappa)x_t + u_t\right](\beta b^\pi_{t-1}+\kappa) - \lambda_{1,t}\gamma_t x_{t-1}(\beta b^\pi_{t-1}+\kappa)- \qquad (47)\\ & - E_t[\lambda_{1,t+1}\beta\gamma_{t+1}((\beta b^\pi_t+\kappa)x_{t+1}+u_{t+1}-b^\pi_t 2x_t)]\\ 0 =& \lambda_{1,t} - \beta E_t\lambda_{1,t+1}(1-\gamma_{t+1}x_t^2) - \beta^2 E_t[((\beta b^\pi_t+\kappa)x_{t+1}+u_{t+1})\,x_{t+1}]- \qquad (48)\\ & \beta^2 E_t\left[\lambda_{1,t+1}\gamma_{t+1}x_t x_{t+1}\right] \end{aligned}$$

Remembering that by Proposition 1 we can concentrate on time-invariant laws of motion for the optimal x, we can rewrite equation (47) as:

$$x_t = \Phi_1\left(b^\pi_{t-1}\right)u_t + \Phi_2\left(s_t\right) \qquad (49)$$

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

$$\begin{aligned} \Phi_1\left(b^\pi_{t-1}\right) &\equiv -\frac{\beta b^\pi_{t-1}+\kappa}{\alpha+\left(\beta b^\pi_{t-1}+\kappa\right)^2} \qquad (50)\\ \Phi_2\left(s_t\right) &\equiv -\frac{1}{\alpha+\left(\beta b^\pi_{t-1}+\kappa\right)^2}\left\{\lambda_{1,t}\gamma_t x_{t-1}(\beta b^\pi_{t-1}+\kappa)\right.\\ & \left. +E_t[\lambda_{1,t+1}\beta\gamma_{t+1}((\beta b^\pi_t+\kappa)x_{t+1}+u_{t+1}-b^\pi_t 2x_t)]\right\} \qquad (51) \end{aligned}$$

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

Plugging (49) into equation (35), we get the following law of motion of b^{π} along any optimal path:

$$b_t^{\pi} = b_{t-1}^{\pi} + \gamma_t x_{t-1} \left[(\beta b_{t-1}^{\pi} + \kappa) \Phi_1 \left(b_{t-1}^{\pi} \right) u_t + u_t - x_{t-1} b_{t-1}^{\pi} \right] + \gamma_t x_{t-1} (\beta b_{t-1}^{\pi} + \kappa) \Phi_2 \left(s_t \right) \tag{52}$$

Using analogous arguments, we get that:

$$b_t^x = b_{t-1}^x + \gamma_t x_{t-1} \left[\Phi_1 \left(b_{t-1}^{\pi} \right) u_t - x_{t-1} b_{t-1}^x \right] + \gamma_t x_{t-1} \Phi_2 \left(s_t \right) \tag{53}$$

Our aim is to rewrite equations (52)-(53) 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 $Y_t \equiv [x_t, x_{t-1}, u_t, \gamma_t,]'$.[12] Hence, we can rewrite (52)-(53) 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:

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

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

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

equations (52)-(53) can be written as:

$$\theta_t = \theta_{t-1} + \gamma_t \mathcal{H} \left(\theta_{t-1}, Y_t \right) + \gamma_t^2 \rho \left(\theta_{t-1}, Y_t \right) \tag{54}$$

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 (54):

$$\frac{d\theta}{d\tau} = h \left(\theta \right) \equiv E \mathcal{H} \left(b^{\pi}, \widehat{Y}_t^2, \widehat{Y}_t^3 \right) \tag{55}$$

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

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.[13] 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 (55) is clearly $\theta = 0$. Moreover it is (locally) stable, since the Jacobian:

$$Dh(\theta) = \begin{pmatrix} -Ex_{t-1}^2(\theta) - b^{\pi}\frac{\partial Ex_{t-1}^2(\theta)}{\partial b^{\pi}} & -b^{\pi}\frac{\partial Ex_{t-1}^2(\theta)}{\partial b^{x}} \\ -b^{x}\frac{\partial Ex_{t-1}^2(\theta)}{\partial b^{\pi}} & -Ex_{t-1}^2(\theta) - b^{x}\frac{\partial Ex_{t-1}^2(\theta)}{\partial b^{x}} \end{pmatrix} \quad (56)$$

has both eigenvalues smaller than zero when evaluated in $\theta = 0$. In the terminology commonly used in the adaptive learning literature, we can say that $\theta = 0$ is the only *E-stable* equilibrium; in Evans and Honkapohja (2001) an equivalence result between E-stability and convergence under learning is derived. However, we cannot directly apply such result, which draws on arguments contained in Benveniste, Métivier, and Priouret (1990), since the state variables' law of motion does not satisfies the required assumptions.[14] However, it turns out that we can adapt their arguments, and prove the following result.

Proposition 3. *Let θ evolve according to* (54)*. If $\overline{\theta}$ is E-stable, then it is locally stable under adaptive learning.*[15]

Proof. See the Appendix. □

Strictly speaking, the above 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; hence, there could exist initial conditions of the states such that there are limiting equilibria which are E-unstable (like for example the RECE). However, numerical investigation show that this is not the case.

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

[14]From a technical point of view, the problem is that 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).

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

5 Constant gain learning

In this section, we analyze the implications of the constant gain learning algorithm. Under this assumption, proving convergence (even locally) is more difficult. Therefore, we look at the ergodic distribution obtained from a Montecarlo experiment. In this case, the economy converges close to REDE if the learning process is slow. However, the learning coefficients b_t^{π} and b_t^{x} converge towards negative values when learning is fast: the long run optimal policy is then to induce expectations cycles in the economy: *ceteris paribus*, when agents observe positive output gap today, they expect deflation and negative output gap tomorrow.

Finally, another natural question is if the optimal policy is a substantial Pareto improvement with respect to policies that drive expectations towards the RECE (Evans and Honkapohja (2006) (EH) policy). In fact, if the welfare difference is small, then following EH might not a bad idea if the monetary authority has strong preference for the long run. Hence, we compare the welfare losses obtained under our optimal policy, and the ones generated by EH. What we find is that the loss induced by EH policy is between 57% and 75% larger than the optimum. We conclude that trying to drive the economy towards RECE may have large costs for the economy.

5.1 The algorithm

Let us reproduce the Lagrangean first-order conditions necessary for an optimum:

$$\begin{aligned} 0 =& -\alpha x_t - \left[(\beta b_{t-1}^{\pi} + \kappa)x_t + u_t\right](\beta b_{t-1}^{\pi} + \kappa) - \lambda_{1,t}\gamma x_{t-1}(\beta b_{t-1}^{\pi} + \kappa) - \\ & - E_t[\lambda_{1,t+1}\beta\gamma((\beta b_t^{\pi} + \kappa)x_{t+1} + u_{t+1} - b_t^{\pi} 2x_t)] \end{aligned} \tag{57}$$

$$0 = \lambda_{1,t} - \beta E_t \lambda_{1,t+1}(1 - \gamma x_t^2) - \beta^2 E_t[((\beta b_t^{\pi} + \kappa)x_{t+1} + u_{t+1})\, x_{t+1}] - \tag{58}$$

$$\beta^2 E_t\left[\lambda_{1,t+1}\gamma x_t x_{t+1}\right] \tag{59}$$

We can solve for $\lambda_{1,t}$ and x_t. The state variables are x_{t-1}, b_{t-1}^{π} and u_t. In order to find a solution, we use a collocation algorithm. This method consists of approximating the control variables as functions of the state variables over few grid points. Typically, one can use Chebychev polynomials for the interpolation, if the functions to be approximated are continuous and smooth (as in the model at hand). Then one needs to find the coefficients of the polynomials that solve the Lagrangean first-order conditions. In our specific case, we generate a three-dimensional grid by choosing the Chebychev zeroes. We approximate $\lambda_{1,t}$ and x_t with Chebychev polynomials[16] and we use tensor product to project the multimensional state space on the policy space. We use quadrature to compute the

[16] In order to generate the approximated policy functions, we use the Miranda-Fackler CompEcon Toolbox.

expectation operators. The coefficients that solve the two equations[17] are found by using a version of the Broyden algorithm for nonlinear equations coded by Michael Reiter. The optimal approximated policy functions are then used to simulate the series.

The benchmark calibration is taken from (Woodford 1999), with $\beta = 0.99$, $\sigma = 0.157$, $\kappa = 0.024$ and $\alpha = 0.04$. The cost-push shocks are drawn from Gaussian distribution with zero mean and variance 0.07. Robustness checks for several of these parameters are also performed. Unless specified, in the simulations we set the initial value for the learning coefficients equal to the values that they would attain under RECE.

5.2 Simulated economies

Figure 2 shows the path for the learning coefficients b_t^π and b_t^x under the benchmark parametrization with constant gain parameter $\gamma = 0.05$. We set initial conditions for learning coefficients equal to the ones in RECE. The economy immediately escapes far away from the RECE. Convergence is relatively fast, with the economy reaching the long run ergodic distribution in around 20000 periods.

However, it is not clear if the economy converges exactly to REDE. In order to check this, we perform a Montecarlo exercise: we draw 10000 realizations of the shock, 100000 periods long, and we simulate the economy starting close to REDE.[18] We then look at the distribution of the learning coefficients b_t^π and b_t^x in the last period of the simulation, which is a good proxy for the ergodic distribution. Figure 3 reports the distributions obtained for different values of the constant gain parameter γ. For small values of the learning parameters, the ergodic distribution is approximately Gaussian with positive mean close to zero. However, when the learning parameter is larger, the distribution is skewed towards negative values. For very large γ, the economy converges towards negative values, which means that in the long run it is optimal to induce cycles in the economy: when there is positive inflation, agents must be induced to believe that there will be deflation in the next period. Note however that standard stochastic approximation results on the convergence of models with constant gain are valid only for γ going to zero (see, among others, Benveniste, Métivier, and Priouret (1990)).

[17]Uniqueness of the solution might be an issue, since the Kuhn-Tucker conditions are only necessary in our setup. However, we experimented with different initial conditions, different interpolation techniques and the solution did not change.

[18]We find that the economy converges close to REDE for any possible initial condition for the learning parameters. In particular, economies starting from, or close to, RECE values always converge close to REDE

5.3 Welfare analysis

Until here, we have shown the qualitative differences between the series of learning coefficients for the optimal policy and the RE-commitment equilibrium. One question is then how much the economy would loose if the monetary authority follows the Evans-Honkapohja (EH) policy rule (i.e., the one that makes private sector learn RECE) instead of the optimal policy derived here. It turns out that the welfare impact is substantial: the welfare loss under EH rule is almost the double of the loss under the optimal policy. Figure 4 and Table 1 illustrate the cumulated welfare loss for different types of learning algorithm (constant and decreasing gain) and for the EH policy. Apart from the learning parameter, all the simulations start from the RECE. All the optimal policy losses are lower than EH, obviously. However the difference is large: in the long run, EH policy has a loss of .44, while all the learning welfare losses are between .2483 and .2822, which implies the EH loss is between 57% and 75% higher than under optimal policy.

Table 1: Welfare Loss

Decr. Gain	Const. Gain, $\gamma = .01$	Const. Gain, $\gamma = .05$	Const. Gain, $\gamma = .1$	EH
0.2584	0.2483	0.2660	0.2822	0.4417

Welfare loss for different policies and learning algorithms.

Figure 4 show the cumulated loss up to each period. We just show the first 1000 periods, since after that the welfare loss is pretty stable. As it can be seen, the optimal policy outperforms the EH policy by large in every periods under every specification of the learning process.

6 Conclusions

Expectations are crucial for monetary policy conduit. We have shown a case where the expectations' formation mechanism yields unexpected results: an optimally behaving central bank does not drive learning agents to the commitment equilibrium, but instead the economy ends up in the discretionary one. Our result can be interpreted as a word of caution in giving general monetary policy rules. In particular, a Taylor rule that would drive the economy towards the commitment equilibrium under rational expectations would deliver big welfare losses in our economy. Therefore, asking the central bank to force convergence of the economy towards a supposedly Pareto-superior RE equilibrium might be misleading.

A natural question is how general is our result. In Stackelberg games with RE there is a clear Pareto ranking between commitment and discretionary equilibria. How sensitive is this ranking to different assumptions about information sets used to form expectations? This is left for future research.

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* (48), *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 (48), 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 $M_Q \equiv \max_{\theta \in Q}(\beta b^\pi + \kappa)$.[19] 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 $|x_t| < \overline{x}\left(u_t\right)$. Because the stochastic process of u is assumed to be iid, it follows that $\overline{\vartheta}_{t+j}$ is independent of $\overline{x}\left(u_{t+1+i}\right)$ and $\overline{\pi}\left(u_{t+1+i}\right)$, 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:

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

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

where $M_{x,\pi} \equiv E_t \overline{x}\left(u_{t+1+i}\right) \overline{\pi}\left(u_{t+1+i}\right)$ 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:

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

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 $\overline{x}\left(u_t\right)$ to get:

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

Finally, note that inequality (A.3) implies that there exists a C_λ such that (A.1) holds.[20] This completes the proof. □

We can now state and prove the following Proposition.

Proposition 4. *Let θ_t evolve according to* (54)*, 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 $\theta \in Q$:*

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

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).

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

Let's start from $\rho_\pi(\cdot)$; plugging equation (51) 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:

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

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:

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

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

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

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:

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

Note that the assumptions on the function ϕ imply that:

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

where the function R, for all θ_k and θ_{k+1} has the upper bound[21]

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

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

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

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

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

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

where:

$$\begin{aligned}
\epsilon_k^{(1)}(\phi) &\equiv \gamma_{k+1}\phi'(\theta_k)\, x_k \left[(\beta b_k^\pi + \kappa)\Phi_1(b_k^\pi)\, u_{k+1} + u_{k+1}, \Phi_1(b_k^\pi)\, u_{k+1}\right]' && \text{(A.15)}\\
\epsilon_k^{(2)}(\phi) &\equiv \gamma_{k+1}\phi'(\theta_k)\, x_k^2 (\theta_k - \theta_{k-1}) && \text{(A.16)}\\
\epsilon_k^{(3)}(\phi) &\equiv (\gamma_k - \gamma_{k+1})\, \phi'(\theta_{k-1})\, x_k^2 \theta_{k-1} && \text{(A.17)}\\
\epsilon_k^{(4)}(\phi) &\equiv \gamma_{k+1}\phi'(\theta_k)\, \theta_k \Phi_1^2(b_k^\pi)\left(\sigma_u^2 - u_{k+1}^2\right) && \text{(A.18)}\\
\epsilon_k^{(5)}(\phi) &\equiv -\gamma_{k+1}\phi'(\theta_k)\, \theta_k \left(\Phi_2^2(s_{k+1}) + 2\Phi_2(s_{k+1})\, \Phi_1(b_k^\pi)\, u_{k+1}\right) && \text{(A.19)}\\
\epsilon_k^{(6)}(\phi) &\equiv \gamma_{k+1}^2 \rho(\theta_k, Y_{k+1}) + R(\phi, \theta_k, \theta_{k+1}) && \text{(A.20)}\\
\eta_{n,r}(\phi) &\equiv -\gamma_{r+1}\phi'(\theta_r)\, x_r^2 \theta_r + \gamma_n \phi'(\theta_{n-1})\, x_n^2 \theta_{n-1} && \text{(A.21)}
\end{aligned}$$

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).[22] 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:*

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

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

Proof. Let's define:

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

and:

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

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

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

zero mean iid shock, which implies that $\overline{Z}_k$ is a martingale difference with respect to F_k; (ii) the following inequality holds:

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

and (iii) the fact that $\overline{Z}_k$ is a martingale difference with respect to F_k implies that[23]:

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

Finally, (iv) we note that:

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

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:

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

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

hence proving the first part of the Lemma; note that we again used the fact that $\phi'\left(\theta\right)$ 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.[24] □

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

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

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

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

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

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

$$I\left(k+1\leq\tau\right)\theta_{k}\Phi_{1}^{2}\left(b_{k}^{\pi}\right)\left(\sigma_{u}^{2}-u_{k+1}^{2}\right) \tag{A.31}$$

is a martingale difference with respect to F_k. □

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

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

Proof. First of all, let's define:

$$D_{n}\equiv\sum_{k=0}^{n-1}I\left(k+1\leq\tau\right)\epsilon_{k}^{5}\left(\phi\right) \tag{A.33}$$

and note that:

$$\begin{aligned}\sup_{n\leq m}I\left(n\leq\tau\right)\left|\sum_{k=0}^{n-1}\epsilon_{k}^{5}\left(\phi\right)\right|\leq\sup_{n\leq m}\left|D_{n}\right| &\leq \sup_{n\leq m}\sum_{k=0}^{n-1}I\left(k+1\leq\tau\right)\left|\epsilon_{k}^{5}\left(\phi\right)\right| &\text{(A.34)}\\ &\leq \sum_{k=0}^{m-1}I\left(k+1\leq\tau\right)\left|\epsilon_{k}^{5}\left(\phi\right)\right| &\text{(A.35)}\end{aligned}$$

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

$$\begin{aligned}\left|\epsilon_{k}^{5}\left(\phi\right)\right| &= \left|\gamma_{k+1}\phi'\left(\theta_{k}\right)\theta_{k}\left(\Phi_{2}^{2}\left(s_{k+1}\right)+2\Phi_{2}\left(s_{k+1}\right)\Phi_{1}\left(b_{k}^{\pi}\right)u_{k+1}\right)\right| &\text{(A.36)}\\ &\leq \left|\gamma_{k+1}\phi'\left(\theta_{k}\right)\theta_{k}\left(\gamma_{k+1}^{2}\left(\frac{\Phi_{2}\left(s_{k+1}\right)}{\gamma_{k+1}}\right)^{2}+2\gamma_{k+1}\frac{\Phi_{2}\left(s_{k+1}\right)}{\gamma_{k+1}}\Phi_{1}\left(b_{k}^{\pi}\right)u_{k+1}\right)\right| &\text{(A.37)}\\ &\leq \left|\gamma_{k+1}^{2}\phi'\left(\theta_{k}\right)\theta_{k}\left(\left(\frac{\Phi_{2}\left(s_{k+1}\right)}{\gamma_{k+1}}\right)^{2}+2\frac{\Phi_{2}\left(s_{k+1}\right)}{\gamma_{k+1}}\Phi_{1}\left(b_{k}^{\pi}\right)u_{k+1}\right)\right| &\text{(A.38)}\\ &\leq \left|\gamma_{k+1}^{2}\phi'\left(\theta_{k}\right)\theta_{k}\widetilde{A}_{5}\left(1+\left|Y_{k+1}\right|^{\overline{q}_{5}}\right)\right| &\text{(A.39)}\end{aligned}$$

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

$$\begin{aligned}&E\left\{\sup_{n\leq m}I\left(n\leq\tau\right)\left|\sum_{k=0}^{n-1}\epsilon_{k}^{5}\left(\phi\right)\right|\right\}^{2}\leq E\left\{\sum_{k=0}^{m-1}I\left(k+1\leq\tau\right)\left|\epsilon_{k}^{5}\left(\phi\right)\right|\right\}^{2} &\text{(A.40)}\\ &\leq \left(\sum_{k=0}^{m-1}\gamma_{k+1}^{2}\right)\left(\sum_{k=0}^{m-1}\gamma_{k+1}^{2}E\left\{I\left(k+1\leq\tau\right)\left|\phi'\left(\theta_{k}\right)\theta_{k}\widetilde{A}_{5}\left(1+\left|Y_{k+1}\right|^{\overline{q}_{5}}\right)\right|^{2}\right\}\right) &\text{(A.41)}\\ &\leq A_{5}\left(1+|y|^{q_{5}}\right)\left(\sum_{k=0}^{m-1}\gamma_{k+1}^{2}\right)^{2} &\text{(A.42)}\end{aligned}$$

□

We can now state and prove our main result.

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

$$E_{y,a}\left\{\sup_{n\leq m} I\left(n\leq\tau\right)\left|\sum_{k=0}^{n-1}\epsilon_k\left(\phi\right)\right|\right\}^2 \leq A\left(1+|y|^q\right)\left(1+\sum_{k=0}^{m-1}\gamma_{k+1}^2\right)\sum_{k=0}^{m-1}\gamma_{k+1}^2 \tag{A.43}$$

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 $\epsilon\left(\phi\right)$ 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.43) is bounded above by expressions of the form:

$$A_i\left(1+|y|^{q_i}\right)\sum_{k=0}^{m-1}\gamma_{k+1}^2 \ \text{ or }\ A_i\left(1+|y|^{q_i}\right)\left(\sum_{k=0}^{m-1}\gamma_{k+1}^2\right)^2 \tag{A.44}$$

By the Cauchy-Schwarz inequality, we have that:

$$\left(\sum_{k=0}^{m-1}\gamma_{k+1}^2\right)^2 \leq \left(\sum_{k=0}^{m-1}\gamma_{k+1}^2\right)\left(\sum_{k=0}^{m-1}\gamma_{k+1}^2\right) \tag{A.45}$$

which implies that the inequality (A.43) 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 $\epsilon\left(\phi\right)$; 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*. □

References

Barro, R. J., and D. B. Gordon (1983): "Rules, discretion and reputation in a model of monetary policy," *Journal of Monetary Economics*, 12(1), 101–121.

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

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

Chari, V. V., and P. J. Kehoe (1990): "Sustainable Plans," *Journal of Political Economy*, 98(4), 783–802.

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.

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.

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.

Haldane, A. G., and V. Read (2000): "Monetary policy surprises and the yield curve," Bank of England working papers 106, Bank of England.

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.

Kydland, F. E., and E. C. Prescott (1977): "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85(3), 473–91.

Levin, A., V. Wieland, and J. C. Williams (1999): *Monetary Policy Rules*chap. Robustness of Simple Monetary Policy Rules under Model Uncertainty, pp. 263–299. Chicago: University of Chicago Press.

Marcet, A., and T. J. Sargent (1989): "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," *Journal of Economic Theory*, 48(2), 337–368.

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.

Molnar, K., and S. Santoro (2010): "Optimal Monetary Policy when Agents are Learning," CESifo Working Paper Series 3072, CESifo Group Munich.

Orphanides, A., and J. C. Williams (2008): "Learning, expectations formation, and the pitfalls of optimal control monetary policy," *Journal of Monetary Economics*, 55(Supplemen), S80–S96.

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.

Sargent, T. J. (1999): *The Conquest of American Inflation.* Princeton: Princeton University Press.

Thornton, D. L. (2003): "Monetary policy transparency: transparent about what?," *Manchester School*, 71(5), 478–497.

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.

Figures

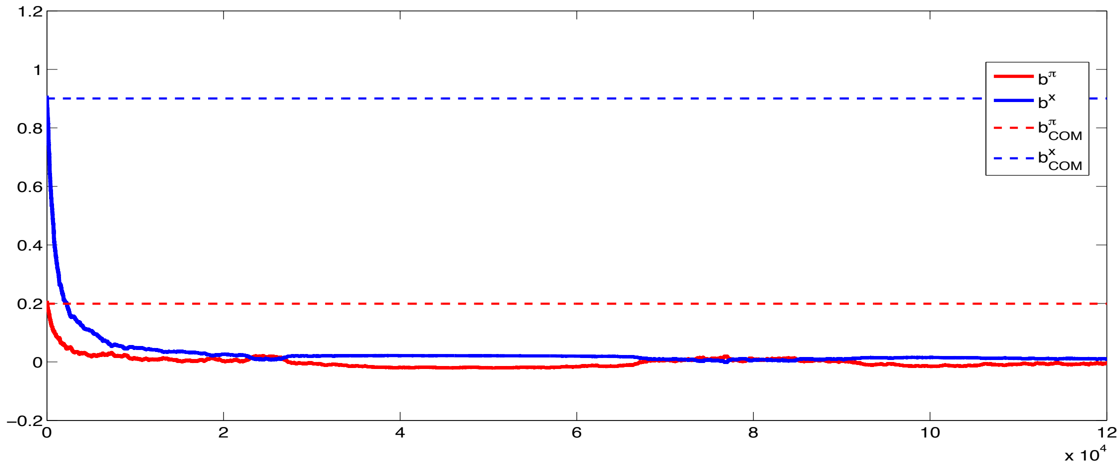

Figure 2: Dynamics of b^{π} and b^{x} under constant gain, benchmark parameterization, $\gamma = .05$

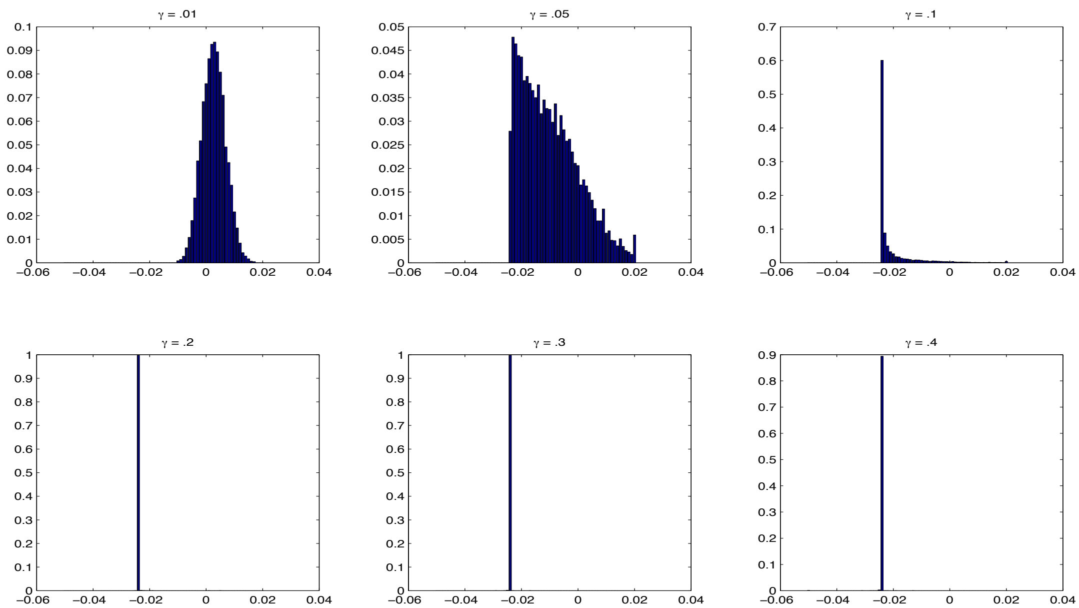

Figure 3: Ergodic distribution (after 10000 draws of 100000 periods), constant gain learning

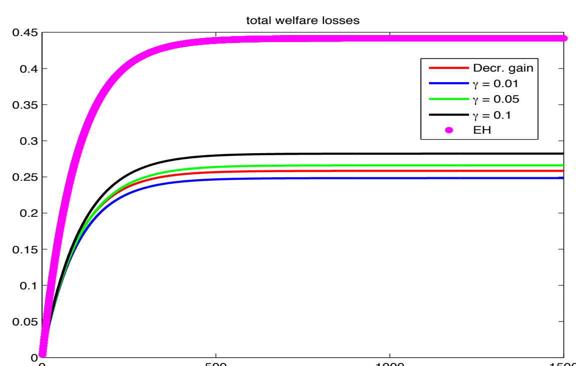

Figure 4: Welfare loss, cumulated