# Error learning behaviour and stability revisited

Domenico Colucci, Vincenzo Valori
DIMAD - Università di Firenze

February 2001
#3-01

**Abstract**

We study the implications of error learning behaviour on the global dynamic properties of stationary equilibria in discrete time deterministic models under bounded rationality. We assume agents' ability to learn from the past performance of their expectations formation mechanism, so that the mechanism itself is made endogenous. We determine sufficient conditions under which this type of error learning behaviour enhances the stability properties of the economy. Also, we show that the set of error learning rules compatible with these conditions is not small in a topological sense and that this set can be used to approximate, with arbitrary precision, alternative learning rules that have been considered in the literature. We give examples drawn from various contexts, and we show how the analysis carried out here relates to e.g. fading memory learning and recursive least squares.

*Journal of Economic Literature* Classification Numbers: C62, D83, D84.

*Key Words:* bounded rationality; convergence to rational expectations; learning dynamics; heterogeneous agents; adaptive expectations; least squares learning.

# 1 Introduction

In this paper we try to develop a context-free, fairly general treatment of the effects on stability of error learning. This problem is all but new in the economic-theoretic literature; in particular, we refer throughout this work to Gerard Fuchs's paper [17] whose language and (to a certain extent) notation we resume here. Several more recent papers deal with this problem in a variety of contexts: these include for instance Evans-Ramey [15, 16], Balasko-Royer [2], Brock-Hommes [8], Arthur *et al.* [1] and Grandmont [20]. It is crucially no less an issue of how one ought to think of stability than it is one of properly defining and modelling error learning. In fact, in our view, the issue of (in)stability of error learning *per se* faces the risk of being nonsensical (or intractable) unless it is analyzed in very simple cases and in a local perspective (Grandmont [20] exhausted most of what was to be said on the matter). Instead, we take a perspective that highlights the effects on stability of the agents' prerogative to revise their expectation mechanism itself according to past prediction performances, with respect to a situation of fixed expectations. In particular we introduce a concept of "stability augmenting" error learning rule, following Fuchs [17], that can be roughly described as follows. For an error learning rule (ELR) i.e. a way of updating the way predictions are formed, to enhance stability we require that 1) trajectories compatible with stability (i.e. converging to a perfect foresight steady state) with fixed expectations are themselves converging under error learning, and that 2) some trajectories, that are not converging with fixed expectations, converge with error learning.

It is clear that the results one is bound to find are strictly dependent on the assumptions regarding the set of ELRs one wants to study. For instance Fuchs [17] studied a particular set of rules in which agents are supposed to consider the occurrence of an exact prediction as a definitive sign that their "vision of the world" is correct and therefore they would not revise it any more. Under such circumstances Fuchs concluded that error learning is in general not stabilizing, in the sense that it actually augments stability in a negligible set of cases (from a topological point of view). In this paper, instead, we study a different set of learning rules, that correspond to a more prudent behavioral attitude: we link agents' willingness to update their expectation schemes (in the presence of self-fulfilled predictions) to their perception of the likelihood of external shocks and to their assessment about the volatility of the relevant economic variables; therefore it takes some time for the agents to become used to a situation of steady environment and this gradual process of building a "habit for stationarity" is precisely what we try to capture through the analytic setup of this paper. Evidence coming from various recent experimental papers (e.g. Slonim [34]) can be invoked as a backup for this idea.[1] Under these conditions we are able to prove that the whole set of learning rules we define is stability-augmenting. Also, we check that this set is not "small" with respect to the entire set of continuously differentiable functions that one can think of (again, we intend "smallness" in a topological sense). Finally, we show that given any Fuchs-style ELR, we can approximate it with some learning rule from the set analysed in this paper.[2]

The paper is organised as follows: Section 2 introduces the notation and all the relevant concepts and presents a rather general result (Theorem 6) on global dynamics which we use extensively throughout the rest of the paper. Section 3 contains the other above mentioned analytical results. Subsection 3.1 discusses the technical issues underlying the choice of an appropriate topology for the results about genericity and convergence. The last section provides examples drawn with a variety of contexts and involves e.g. fading memory learning and recursive least squares. The Appendix contains some of the (less interesting) proofs.

# 2 A general result

We start from a very general class of discrete-time models in which the relation between a vector of variables at time $s$, $\mathbf{x}_s \in \mathbb{X} \subset \mathbb{R}^n$ and the vector $\mathbf{x}_t^e \in \mathbb{X}$ of agents' expectations on the time-$t$ value of the state $\mathbf{x}$, with $s = t - k$, and $k$ an integer, is determined by a temporary equilibrium equation

$$Z\left(\mathbf{x}_s, \mathbf{x}_t^e, \mathbf{x}_{t-1}^e, \ldots, \mathbf{x}_{t-P}^e\right) = \mathbf{0} \tag{1}$$

where $Z : \mathbb{R}^{(P+2)n} \to \mathbb{R}^n$, can be thought of as some kind of market clearing condition. Equation (1) is consistent with a large variety of models;[3] for

[1] Unfortunately in economics, the consequences of experimental evidence for theoretical modelling can almost invariably be argued about.

[2] In other words, given an ELR *à la Fuchs*, we can construct an appropriate sequence of learning rules converging (in a sense to be made precise) to it.

[3] Notice that it is sufficient to consider $\mathbf{x}_t^e \in \mathbb{R}^{n\times h}$ and $Z : \mathbb{R}^{n[1+(P+1)h]} \to \mathbb{R}^n$ to encompass the case of (finitely many) heterogeneous agents. Indeed, the temporary equilibrium

instance a simple overlapping generations economy would have $n = 1$ (one good), $s = t - 1$ (forward looking behaviour) and $P = 1$ (two-periods-living agents), whereas for a cobweb model $n = 1$, $s = t$ and $P = 0$ would be appropriate. Notice that with $k > 0$ (1) implies a forward-looking feature. A local version of the model (in terms of a *perfect foresight stationary equilibrium* $\mathbf{x}^*$ such that $Z(\mathbf{x}^*, \mathbf{x}^*, \ldots, \mathbf{x}^*) = \mathbf{0}$) can be written, under the assumptions of the Implicit Function Theorem on the regularity of the map $Z$:

$$\mathbf{x}_s = F(\mathbf{x}_t^e, \ldots, \mathbf{x}_{t-P}^e) \tag{2}$$

We consider economic models in which agents form their expectations relying on a vector of past observations and predictions about the state variables; this is done by a vector valued expectation function $\mathbf{e}(\cdot)$ which is selected within a set of predictors parametrized by a vector $\boldsymbol{\alpha} \in \mathbb{R}^m$

$$\mathbf{x}_{t+1}^e = \mathbf{e}_{\boldsymbol{\alpha}}\left(\mathbf{x}_s, \ldots, \mathbf{x}_{s-T}, \mathbf{x}_t^e, \ldots, \mathbf{x}_{t-R}^e\right) \tag{3}$$

Remark that the first argument of the expectation function is $\mathbf{x}_s$: this is because, due to (2), more recent observations are not available.[4]

**Assumption 1 (A1)** *The expectation function in (3) satisfies the following rationality requirement*

$$F(\mathbf{x}^*, \ldots, \mathbf{x}^*) = \mathbf{x}^* \Rightarrow \mathbf{e}_{\boldsymbol{\alpha}}(\mathbf{x}^*, \ldots, \mathbf{x}^*) = \mathbf{x}^*$$

Under the above assumption agents are supposed to be able to detect a stationary point when they have enough information. It is a standard requirement (see Grandmont [20] for an example).

Notice that the choice of the actual predictor is made by each agent according to the value of $\boldsymbol{\alpha}$. Plugging $\mathbf{x}_s = F(\mathbf{x}_t^e, \ldots, \mathbf{x}_{t-P}^e)$ into (3) we get

$$\begin{aligned} \mathbf{x}_{t+1}^e &= \mathbf{e}_{\boldsymbol{\alpha}}\left(F(\mathbf{x}_t^e, \ldots, \mathbf{x}_{t-P}^e), \ldots, F(\mathbf{x}_{t-T}^e, \ldots, \mathbf{x}_{t-P-T}^e), \mathbf{x}_t^e, \ldots, \mathbf{x}_{t-R}^e\right) = \\ &= \bar{\mathbf{e}}_{\boldsymbol{\alpha}}\left(\mathbf{x}_t^e, \ldots, \mathbf{x}_{t-K}^e\right) \qquad \text{with } K = \max\{P + T, R\} \end{aligned} \tag{4}$$

map can be thought of as one in which a vector of state variables, $\mathbf{x}_s$, is a function of the past $P + 1$ predictions of each of the $h$ agents of the economy. Setting $h = 1$ we are left with a representative agent.

[4]Remark that (3) is compatible both with the case in which traders use a rolling window of data of length $T$ and the case of unbounded memory (provided that it can be written in a recursive form); see Section 4 for examples.

Consider now the case in which $\boldsymbol{\alpha}$ is, in turn, endogenously determined according to a general law of the form

$$\boldsymbol{\alpha}_{t+1} = \mathbf{g}\left(\mathbf{x}_s, \ldots, \mathbf{x}_{s-T}, \mathbf{x}_t^e, \ldots, \mathbf{x}_{t-R}^e, \boldsymbol{\alpha}_t\right) \tag{5}$$

While we try to keep the formalism very general we can think of $\boldsymbol{\alpha}$ as a function of the performances of past predictions. In particular what we have in mind is some form of *error learning behaviour* in the sense of Fuchs [17]. Alternative ways of modelling forecasts formation through the selection of alternative expectation functions can be found in the recent literature: Evans and Ramey [16], Brock and Hommes [8] are two examples.

We analyse the dynamical system obtained from (2), (3) and (5) substituting for $\mathbf{x}_s, \mathbf{x}_{s-1}, \ldots, \mathbf{x}_{s-T}$ in (3) and (5):

$$\begin{cases} \mathbf{x}_{t+1}^e = \bar{\mathbf{e}}\left(\mathbf{x}_t^e, \ldots, \mathbf{x}_{t-K}^e, \boldsymbol{\alpha}_t\right) \\ \boldsymbol{\alpha}_{t+1} = \bar{\mathbf{g}}\left(\mathbf{x}_t^e, \ldots, \mathbf{x}_{t-K}^e, \boldsymbol{\alpha}_t\right) \end{cases} \tag{6}$$

It is convenient to rewrite (6) in a more compact form. Introducing the auxiliary variables

$$\begin{aligned} \mathbf{y}_t^1 &= \mathbf{x}_{t-1}^e \\ &\vdots \\ \mathbf{y}_t^K &= \mathbf{y}_{t-1}^{K-1} \end{aligned}$$

(6) becomes

$$\begin{cases} \mathbf{y}_{t+1}^K = \mathbf{y}_t^{K-1} \\ \vdots \\ \mathbf{y}_{t+1}^1 = \mathbf{x}_t^e \\ \mathbf{x}_{t+1}^e = \bar{\mathbf{e}}\left(\mathbf{x}_t^e, \mathbf{y}_t^1, \ldots, \mathbf{y}_t^K, \boldsymbol{\alpha}_t\right) \\ \boldsymbol{\alpha}_{t+1} = \bar{\mathbf{g}}\left(\mathbf{x}_t^e, \mathbf{y}_t^1, \ldots, \mathbf{y}_t^K, \boldsymbol{\alpha}_t\right). \end{cases}$$

Then, defining

$$\mathbf{X}_t = \left(\mathbf{x}_t^e, \mathbf{y}_t^1, \ldots, \mathbf{y}_t^K\right)$$

we end up with

$$\begin{cases} \mathbf{X}_{t+1}^e = \mathbf{E}\left(\mathbf{X}_t^e, \boldsymbol{\alpha}_t\right) \\ \boldsymbol{\alpha}_{t+1} = \mathbf{G}\left(\mathbf{X}_t^e, \boldsymbol{\alpha}_t\right) \end{cases} \tag{7}$$

We will call the $\mathbf{G}(\cdot)$ function in (7) an *error learning rule*[5].

Following this type of notation we can rewrite the 'static' model (4) as follows:

$$\mathbf{X}_{t+1}^{e} = \mathbf{E}_{\boldsymbol{\alpha}}\left(\mathbf{X}_{t}^{e}\right) \tag{8}$$

In this paper we aim at comparing the stability properties of the fixed points for the two models (7) and (8). We will follow Fuchs [17] quite closely in the way this stability comparison is done. This is described precisely in the following definitions.

**Definition 2** *Given the fixed point $\mathbf{x}^* = F(\mathbf{x}^*)$ and $\mathbf{X}^* = (\mathbf{x}^*, \ldots, \mathbf{x}^*)$; let $A(\mathbf{X}^*)$ be the set of parameter values, $\boldsymbol{\alpha}$, for which $\mathbf{X}^*$ is asymptotically stable and hyperbolic for the equation (8).*

---

[5] Observe that, while we have preferred to make explicit the dynamic of the system with respect to the expected value $x^e$ - leaving the state variable $x$ hidden, although univocally determined, in the underlying temporary equilibrium map - this is not the only possible choice. Wishing to follow the opposite way (as Grandmont in [20]) the model should be slightly changed as follows.

Given a temporary equilibrium equation of the form

$$Z\left(\mathbf{x}_t, \ldots, \mathbf{x}_{t-P}, \mathbf{x}_s^e, \ldots, \mathbf{x}_{s-R}^e\right) = \mathbf{0} \tag{FN1}$$

let's consider traders who perform their forecasts, relying on a vector of past observations, through an expectation function which we suppose to be parametrized by a vector $\boldsymbol{\alpha}$

$$\mathbf{x}_s^e = \mathbf{e}_{\boldsymbol{\alpha}}\left(\mathbf{x}_t, \ldots, \mathbf{x}_{t-T}\right) \tag{FN2}$$

Define $K = \max\{P, T+R\}$; by substitution of (FN2) in (FN1) the temporary equilibrium map becomes

$$Z\left(\mathbf{x}_t, \ldots, \mathbf{x}_{t-K}, \boldsymbol{\alpha}\right) = \mathbf{0}$$

Now suppose that, in a neighborhood of the *perfect foresight stationary equilibrium* $\mathbf{x}^*$, the assumptions of the Implicit Function Theorem on the regularity of the map $Z$ hold so that a local version of the model can be written

$$\mathbf{x}_t = F(\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-K}, \boldsymbol{\alpha})$$

The rest of the story remains unchanged, leading to the system

$$\begin{cases} \mathbf{X}_{t+1} = \mathbf{E}\left(\mathbf{X}_t, \boldsymbol{\alpha}_t\right) \\ \boldsymbol{\alpha}_{t+1} = \mathbf{G}\left(\mathbf{X}_t, \boldsymbol{\alpha}_t\right) \end{cases} \tag{FN3}$$

which is analogous, from an algebraic point of view, to 7. All the results in this paper could be referred to (FN3).

**Definition 3** *An error learning rule for the model (7) is said to be* ***stability-augmenting*** *with respect to the model (8) if there is a neighbourhood* $\mathbf{U}$ *of* $\mathbf{X}^*$ *such that:*

1. *trajectories with initial conditions* $(\mathbf{X}_0, \boldsymbol{\alpha}_0)$ *with* $\mathbf{X}_0 \in \mathbf{U}$ *and* $\boldsymbol{\alpha}_0 \in A(\mathbf{X}^*)$ *converge to* $(\mathbf{X}^*, \boldsymbol{\alpha}^*)$ *with* $\boldsymbol{\alpha}^* \in A(\mathbf{X}^*)$*;*

2. *there are trajectories starting at* $(\mathbf{X}_0, \boldsymbol{\alpha}_0)$ *with* $\mathbf{X}_0 \in \mathbf{U}$ *and* $\boldsymbol{\alpha}_0 \notin A(\mathbf{X}^*)$ *that converge to* $(\mathbf{X}^*, \boldsymbol{\alpha}^*)$ *with* $\boldsymbol{\alpha}^* \in A(\mathbf{X}^*)$*.*

Basically *1.* says that with any $\boldsymbol{\alpha}_0 \in A(\mathbf{X}^*)$, the equilibrium remains attracting as in the static model (so that selecting predictors according to error learning should not make things worse); *2.* renders the idea that learning can turn a situation of instability of the equilibrium into one compatible with convergence to self-fulfilling expectations (so that error learning should actually help stability at least in some cases).

As it turns out, the fate of the stability issue depends crucially on the assumptions regarding the error learning rules that are deemed to be important from the economic point of view. We now define the set of learning rules of interest. For an interpretation of the key assumptions we are making and for a discussion about existing alternatives from the literature see the next section.

**Assumption 4 (A2)** *Given the model (7) and* $\mathbf{X}^*$ *as in Definition 2, let the error learning rules of interest be the set of continuously differentiable functions* $\mathbf{G} : \mathbb{R}^{nT} \times \mathbb{R}^m \to \mathbb{R}^m$ *such that there is* $\boldsymbol{\alpha}^*$ *and* $\gamma \in (0,1)$ *such that*

$$\|\boldsymbol{\alpha}^* - \mathbf{G}(\mathbf{X}^*, \boldsymbol{\alpha}_t)\| \leq \gamma \|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_t\|$$

*for all* $\boldsymbol{\alpha}_t$*, where* $\|\cdot\|$ *is the Euclidean norm.*

**Definition 5** *The subset of error learning rules satisfying (A2) that have* $\boldsymbol{\alpha}^* \in A(\mathbf{X}^*)$ *(see Definition 2) is called* $\Pi$*.*

Let us now turn to the general result.

**Theorem 6** *Given the system (7) consider a fixed point* $\mathbf{x}^* = F(\mathbf{x}^*)$ *and* $\mathbf{X}^* = (\mathbf{x}^*, \ldots, \mathbf{x}^*)$*, assume (A1) and let* $\mathbf{G} \in \Pi$*.*

*Then, for all* $\boldsymbol{\alpha} \in \mathbb{R}^m$ *there exists* $\delta > 0$ *(depending on* $\boldsymbol{\alpha}$*) such that*

$$B_\delta(\mathbf{X}^*) \times \{\boldsymbol{\alpha}\} \subset \mathfrak{B}(\mathbf{X}^*, \boldsymbol{\alpha}^*) \tag{9}$$

*where* $B_\delta(\mathbf{X}^*)$ *is a ball of radius* $\delta$ *centred in* $\mathbf{X}^*$ *and* $\mathfrak{B}(\mathbf{X}^*, \boldsymbol{\alpha}^*)$ *is the basin of attraction of* $(\mathbf{X}^*, \boldsymbol{\alpha}^*)$*.*

Before proving this result we need the following lemma.

**Lemma 7** *Under the assumptions of Theorem 6 the couple* $(\mathbf{X}^*, \boldsymbol{\alpha}^*)$ *is an asymptotically stable fixed point for the system (7).*

**Proof.**

- Thanks to (A2) and continuity there is a suitable $\eta$ such that if $\mathbf{X}_t^e \in I_\eta(\mathbf{X}^*)$ there is $\xi_1 \in (\gamma, 1)$ such that $\|\boldsymbol{\alpha}^* - \mathbf{G}(\mathbf{X}^*, \boldsymbol{\alpha}_t)\| \leq \xi_1 \|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_t\|$ for all $\boldsymbol{\alpha}_t$.

- Notice that, because $\boldsymbol{\alpha}^* \in A(\mathbf{X}^*)$, the dynamics for points $(\mathbf{X}_t^e, \boldsymbol{\alpha}^*)$ is such that there is $\lambda \in (0,1)$ and $0 < \eta_1 \leq \eta$ such that

$$\|\mathbf{X}^* - \mathbf{E}(\mathbf{X}_t, \boldsymbol{\alpha}^*)\| < \lambda \|\mathbf{X}^* - \mathbf{X}_t\| \tag{10}$$

  holds for all $\mathbf{X}_t \in I_{\eta_1}(\mathbf{X}^*)$.

- Again, continuity implies that there is $\eta_2 > 0$ and $\xi_2 \in (\lambda, 1)$ for which

$$\|\mathbf{X}^* - \mathbf{E}(\mathbf{X}_t, \boldsymbol{\alpha}_t)\| < \xi_2 \|\mathbf{X}^* - \mathbf{X}_t\|$$

  holds for all $\mathbf{X}_t \in I_{\eta_1}(\mathbf{X}^*)$ and $\boldsymbol{\alpha}_t \in I_{\eta_2}(\boldsymbol{\alpha}^*)$.

- Summing up if $\xi = \max\{\xi_1, \xi_2\}$ and $\delta = \min\{\eta_1, \eta_2\}$, then for all $(\mathbf{X}_t, \boldsymbol{\alpha}_t) \in B_\delta(\mathbf{X}^*, \boldsymbol{\alpha}^*)$ we have

$$\|(\mathbf{X}^*, \boldsymbol{\alpha}^*) - (\mathbf{X}_{t+1}, \boldsymbol{\alpha}_{t+1})\| \leq \xi \|(\mathbf{X}^*, \boldsymbol{\alpha}^*) - (\mathbf{X}_t, \boldsymbol{\alpha}_t)\|$$

**Proof.** [of Theorem 6]

The strategy of this proof is to build, through backward induction, a sequence of neighbourhoods of the stationary state, belonging to its basin of attraction.

Thanks to Lemma 7, there are $\delta_0, \mu_0 > 0$ such that

$$I_0(\mathbf{X}^*, \boldsymbol{\alpha}^*) = \left\{B_{\delta_0}(\mathbf{X}^*) \times B_{\mu_0}(\boldsymbol{\alpha}^*)\right\} \subset \mathfrak{B}(\mathbf{X}^*, \boldsymbol{\alpha}^*)$$

Given $\hat{\boldsymbol{\alpha}}_0 \neq \boldsymbol{\alpha}^*$ such that $\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_0\| \leq \mu_0$ and $\gamma$ satisfying (A2), consider the sequence of vectors

$$\hat{\boldsymbol{\alpha}}_{n+1} = \frac{\hat{\boldsymbol{\alpha}}_n - (1-\gamma)\boldsymbol{\alpha}^*}{\gamma} \tag{11}$$

We have

$$\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_{n+1}\| = \left\|\boldsymbol{\alpha}^* - \frac{\hat{\boldsymbol{\alpha}}_n - (1-\gamma)\boldsymbol{\alpha}^*}{\gamma}\right\| = \left\|\frac{\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_n}{\gamma}\right\| = \frac{1}{\gamma}\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_n\|$$

and therefore $\{\mu_n\} \equiv \{\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_n\|\}$ is monotone increasing and tends to $+\infty$.

Consider now $C_n \equiv \left\{B_{\mu_{n+1}}(\boldsymbol{\alpha}^*) - B_{\mu_n}(\boldsymbol{\alpha}^*)\right\}$. For any $\boldsymbol{\alpha}_t \in \bar{C}_n$ (the closure of $C_n$), thanks to [A2] and (11), we have

$$\begin{aligned}\|\boldsymbol{\alpha}^* - G(\mathbf{X}^*, \boldsymbol{\alpha}_t)\| &= \|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_{t+1}\| < \gamma\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_t\| \leq \gamma\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_{n+1}\| = \\ &= \|\gamma\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_n + (1-\gamma)\boldsymbol{\alpha}^*\| = \|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_n\| = \mu_n\end{aligned}$$

Because the map $G(\cdot)$ is uniformly continuous on $\overline{B_{\delta_0}(\mathbf{X}^*, \boldsymbol{\alpha}) \times C_n}$ we can find $\delta'_{n+1} > 0$ such that

$$\|\boldsymbol{\alpha}^* - G(\mathbf{X}^e_t, \boldsymbol{\alpha}_t)\| < \|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}_n\| = \mu_n \tag{12}$$

for any $(\mathbf{X}^e_t, \boldsymbol{\alpha}_t) \in B_{\delta'_{n+1}}(\mathbf{X}^*) \times \mathbf{C}_n$. Also (using the uniform continuity of $\mathbf{E}(\cdot)$ this time) there is $0 < \delta_{n+1} \leq \delta'_{n+1}$ such that for any $(\mathbf{X}_t, \boldsymbol{\alpha}_t) \in B_{\delta_{n+1}}(\mathbf{X}^*) \times C_n$

$$\left\|\mathbf{X}^e_{t+1} - \mathbf{X}^*\right\| \leq \delta_n \tag{13}$$

Now given $\delta_0, \mu_0 > 0$, thanks to (12) and (13), we can retrieve recursively $\{(\delta_n, \mu_n)\}$, $\{\delta_n\}$ being decreasing and $\{\mu_n\}$ increasing and diverging to $+\infty$, such that

$$(\mathbf{X}_t, \boldsymbol{\alpha}_t) \in \left\{B_{\delta_{n+1}}(\mathbf{X}^*) \times C_n\right\} \Rightarrow (\mathbf{X}_{t+1}, \boldsymbol{\alpha}_{t+1}) \in \left\{B_{\delta_n}(\mathbf{X}^*) \times B_{\mu_n}(\boldsymbol{\alpha}^*)\right\}$$

So given $\boldsymbol{\alpha}$ we can find $n$ such that $\boldsymbol{\alpha} \in C_n$ and, correspondingly, $\delta = \delta_{n+1}$ such that for any $\mathbf{X} \in B_\delta(\mathbf{X}^*)$ the system takes $(\mathbf{X}, \boldsymbol{\alpha})$ to $I_0(\mathbf{X}^*, \boldsymbol{\alpha}^*) = \left\{B_{\delta_0}(\mathbf{X}^*) \times B_{\mu_0}(\boldsymbol{\alpha}^*)\right\} \subset \mathfrak{B}(\mathbf{X}^*, \boldsymbol{\alpha}^*)$, in finite steps (at most $n$). This concludes our proof. ■

Theorem 6 suggests that the choice of a learning rule in the set $\Pi$ will make trajectories converge to the fixed point $(\mathbf{X}^*, \boldsymbol{\alpha}^*)$ for all initial values of the learning variable $\boldsymbol{\alpha}$, provided that the initial value for the state variable is sufficiently near to its steady state level.[6] Also, as it can be inferred from the proof, we have a sort of "eye-shaped" region embedded in the basin of attraction of the steady state. Figure 1 provides a pictorial representation. This fact has some nice consequences that we exploit extensively in the next Section.

---

[6]Remark that initial conditions in the system (7) have to be expressed in terms of expectations, $\mathbf{x}^e_t$. In our model though the expectations functions (3) need to be initialized with values of the state variable, $\mathbf{x}_t$. As a consequence of (A1) and of continuity, if the initial state is close to its perfect-foresight level so will be the initial expectations.

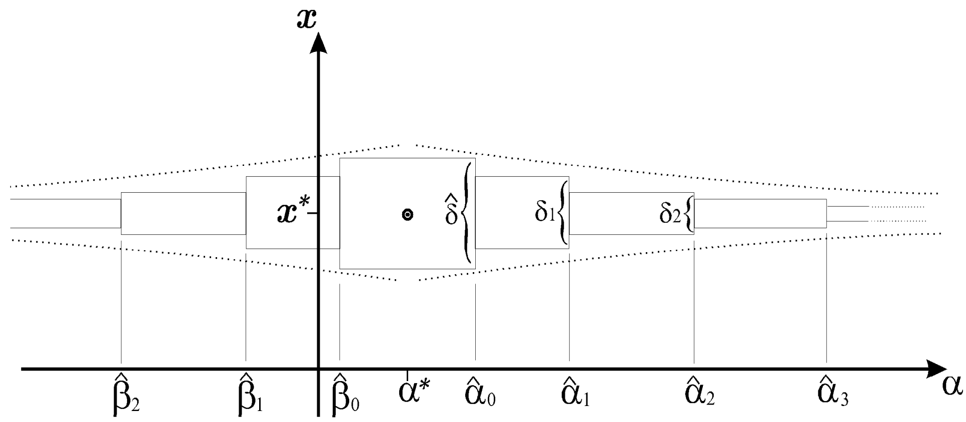


Figure 1:

# 3 Comparative findings

We now show how this result can be used to understand the implications of error learning behaviour in dynamic economic models. This issue has been studied in Fuchs [17] considering a model of economic dynamics whose evolution depends crucially on agent's predictions about future values of some significant variable describing the state of the system. In particular [17] develops on the idea that if more skilled agents give rise to economic system that are more stable, as it is claimed by a classical intuition, one naturally expects equilibrium convergence to occur more often than in models that have a fixed expectations updating rule. This is done in the technical framework of Definition 3.

In details, [17] works with the set of continuously differentiable $ELRs$ defined on a compact[7] such that

$$\boldsymbol{\alpha}_{t+1} = \mathbf{G}\left(\mathbf{X}_t - \mathbf{X}_t^e, \boldsymbol{\alpha}_t\right)$$

with

$$\mathbf{G}\left(\mathbf{0}, \boldsymbol{\alpha}\right) = \boldsymbol{\alpha} \tag{14}$$

The set of these learning rules is labelled $F$.

In this situation we have the following:

[7] This is possible because Fuchs worked with an overlapping generations environment with the prices (that can ber normalized) as the state variable.

**Proposition 8 (Fuchs 79)** *Given the system 7, let $\mathcal{L} \subset F$ be the subset of the stability-augmenting learning rules (according to Definition 3), with respect to the fixed point $\mathbf{X}^*$. Then $\mathcal{L}$ is a closed set with an empty interior (in the $\tau_1$ topology).*

**Proof.** See [17].

This result (a slightly simplified version of the original) states that learning devices of the type described by Fuchs enhance the stability of the system in a negligible set of cases. This conclusions draws heavily on assumption (14), and its effects on the expectation function used by the agents, which is seen as their "vision of things". (14) simply says that, in case of perfect anticipation of the future prices, agents would immediately deduce that their theory is right and would therefore leave it unchanged to formulate forecasts in the next period. The negative conclusion of Proposition 8 appears to imply quite a strong statement which has not received much validation from the empirical and experimental literature in more recent years. Also, in many models used in the literature, assumption (14) does not hold. As we show in the sequel it is possible to reconcile this type of theoretic set-up with (at least part of) the empirical evidence available.

We concentrate on the class of learning rules described in Assumption (A2) and Definition 5, which correspond to a more prudent attitude on the agents' part. The underlying idea is that, in our vision, the internal variable $\boldsymbol{\alpha}_t$ is to be connected to the possibility of external shocks as it is perceived by the agents and to their assessment about the volatility of the economic environment: therefore, upon its first appearance, the occurrence of fulfilled expectations will not be immediately regarded as a situation of stationary equilibrium, either because the possibility of shocks is felt as possible or because it is only considered a fluke. But as time passes and expectations are confirmed to be exact agents will get used to the idea of equilibrium and the possibility of shocks will be perceived as less and less likely. This process is captured in the Assumption (A2), requiring that when $\mathbf{X}_t = \mathbf{X}^*$ the internal variable $\boldsymbol{\alpha}_t$ must tend to an equilibrium value $\boldsymbol{\alpha}^*$ of its own, with a given speed of convergence $\gamma$.

It is straightforward to check that **all** the learning rules that satisfy the assumptions of Theorem 6 (the set $\Pi$) are stability-augmenting:

**Theorem 9** *Given the system (7) consider a fixed point $\mathbf{x}^* = F(\mathbf{x}^*)$ and $\mathbf{X}^* = (\mathbf{x}^*, \ldots, \mathbf{x}^*)$. Then every element of the set $\Pi$ of Definition 5 is stability-augmenting according to Definition 3.*

**Proof.**

If $\boldsymbol{\alpha}_0 \in A(\mathbf{X}^*)$ then Lemma 7 shows that the system will converge to $(\mathbf{X}^*, \boldsymbol{\alpha}^*)$. Conversely suppose $\boldsymbol{\alpha}_0 \notin A(\mathbf{X}^*)$: then by Theorem 6, given $\mathbf{G} \in \Pi$ and $\boldsymbol{\alpha}$, we can find $\delta > 0$ (depending on $\boldsymbol{\alpha}$) such that

$$B_\delta(\mathbf{X}^*) \times \{\boldsymbol{\alpha}\}$$

is in the basin of attraction of (and therefore converges to) $(\mathbf{X}^*, \boldsymbol{\alpha}^*)$.

■

Hence, if we substitute the assumption (14) of Fuchs with (A2), the negative results in [17] is completely reversed.

In fact we can say even more. As we anticipated the subset of learning rules we are considering is not a "thin set", in a topological sense. This is what we are going to show next, in Theorem 15.

But first we need to define a topology for the set of functions we are interested in. One possibility is to restrict to bounded functions or to consider a compact domain, and so resort to the $\tau_1$ topology of $C^1$-uniform convergence. This is clearly a heavy requirement in terms of the underlying economic interpretation. Hence, we will stick to a higher level of generality.

Let $\mathcal{F}$ be the set of $C^1$ functions defined on $\mathbb{R}^s$ taking values in $\mathbb{R}^p$. Consider, on this set, the following distance:

**Definition 10** *Consider a "telescopic" sequence of non-empty compact sets $K_i \subset K_{i+1}$ such that $\bigcup_i K_i = \mathbb{R}^s$. On every compact $K_i$ define*

$$\max_{x \in K_i} (\|f(x)\| + \|Df(x)\|)$$

*where $\|f(x)\|$ is the euclidean norm in $\mathbb{R}^n$ and $\|Df(x)\|$ is the norm the of the Jacobian $Df(x)$ defined by*

$$\|Df(x)\| = \max_{\|v\|=1} \|Df(x)\, v\|$$

*Let, $\tau_p : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$, be the following function:*

$$\tau_p(f,g) = \lim_{i \to \infty} \frac{\max_{x \in K_i} (\|f(x) - g(x)\| + \|Df(x) - Dg(x)\|)}{1 + \max_{x \in K_i} (\|f(x) - g(x)\| + \|Df(x) - Dg(x)\|)}$$

**Proposition 11** *The function $\tau_p(f,g)$ is a metric.*

**Proof.**

Clearly $\tau_p(f,g) \geq 0$ and $f \neq g$ implies that there is $k_i$ such that

$$\max_{x \in K_i} (\|f(x) - g(x)\| + \|Df(x) - Dg(x)\|) > 0$$

implying $\tau_p(f,g) > 0$. Also, trivially, $\tau_p(f,g) = \tau_p(g,f)$.

Finally $\tau_p(f,g) \leq \tau_p(f,h) + \tau_p(h,g)$. Indeed, given a compact $K_i$, let $\tau_{1,i}$ be the metric of the $C^1$-uniform convergence for functions defined on $K_i$. Because $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi(t) = \frac{t}{1+t}$ is increasing, and thanks to the triangular inequality

$$\begin{aligned} \frac{\tau_{1,i}(f,g)}{1+\tau_{1,i}(f,g)} &\leq \frac{\tau_{1,i}(f,h)+\tau_{1,i}(h,g)}{1+\tau_{1,i}(f,h)+\tau_{1,i}(h,g)} = \\ &= \frac{\tau_{1,i}(f,h)}{1+\tau_{1,i}(f,h)+\tau_{1,i}(h,g)} + \frac{\tau_{1,i}(h,g)}{1+\tau_{1,i}(f,h)+\tau_{1,i}(h,g)} \leq \\ &\leq \frac{\tau_{1,i}(f,h)}{1+\tau_{1,i}(f,h)} + \frac{\tau_{1,i}(h,g)}{1+\tau_{1,i}(h,g)} \end{aligned}$$

so that taking the limit for $i \to \infty$ we get the desired inequality. ■

**Remark 1** *Note that $\tau_p(f,g) \leq 1$ for any $f,g$ belonging to $\mathcal{F}$. The metric $\tau_p$ on the set $\mathcal{F}$ does not come from any norm since, for any $\lambda \neq 0$*

$$|\lambda| \tau_p(f,0) \neq \tau_p(\lambda f,0)$$

**Proposition 12** *The distance $\tau_p$ does not depend on the choice of the telescopic sequence of compact sets.*

**Proof.** In the Appendix.

**Remark 2** *Observe that*

$$\|f(x) - g(x)\| + \|Df(x) - Dg(x)\| < \varepsilon \qquad \text{for any } x \in \mathbb{R}^m$$

*implies $\tau_p(f,g) < \varepsilon$ and that, on the contrary, if $\tau_p(f,g) < \sigma$ then for all $x \in \mathbb{R}^m$*

$$\|f(x) - g(x)\| + \|Df(x) - Dg(x)\| < \frac{\sigma}{1-\sigma}$$

*which is smaller than $2\sigma$ for $\sigma < \frac{1}{2}$. In other words, two functions are near for the $\tau_p$ metric if and only if they are pointwise $\varepsilon$-near everywhere.*

**Lemma 13** *Let $f \in \mathcal{F}$ a contraction of constant $\gamma$. For any $\varepsilon$ such that $\gamma + \varepsilon < 1$, there is $\delta_\varepsilon$ such that*

$$\tau_p(f, g) < \delta_\varepsilon \Longrightarrow g \text{ is a contraction of constant } \beta < \gamma + \varepsilon$$

**Proof.** In the Appendix.

**Lemma 14** *Let $f \in \mathcal{F}$ be a contraction of constant $\gamma$, $z^* = f(z^*)$ and $B_\eta(z^*)$ an $\eta$-ball centred in $z^*$: then $f(B_\eta(z^*)) \subseteq B_{\eta\gamma}(z^*)$.*

**Proof.** By definition, for all $z \in B_\eta(z^*)$ we have

$$\|f(z) - z^*\| = \|f(z) - f(z^*)\| \leq \gamma \|z - z^*\| \leq \gamma \max_{y \in B_\eta(z^*)} \|y - z^*\| = \eta\gamma$$

■

We are now ready to prove that the subset of learning rules we are considering is not a "thin set", in the sense that it contains open sets.

**Theorem 15** *Given the system (7) consider a fixed point $\mathbf{x}^* = F(\mathbf{x}^*)$ and $\mathbf{X}^* = (\mathbf{x}^*, \ldots, \mathbf{x}^*)$. Define the set*

$$\Gamma = \left\{\mathbf{G} : \mathbb{R}^{nT} \times \mathbb{R}^m \to \mathbb{R}^m,\ \mathbf{G} \in \mathbf{C}^1\right\}$$

*of the continuously differentiable learning rules, endowed with the $\tau_p$ topology. Then $\Pi$, as a subset of $\Gamma$, has non-empty interior.*

**Proof.**

To show that the set is not empty consider

$$\hat{G}(\mathbf{X}_t^e, \boldsymbol{\alpha}_t) = \boldsymbol{\alpha}^* + \gamma_1 (\boldsymbol{\alpha}_t - \boldsymbol{\alpha}^*)$$

with $\boldsymbol{\alpha}^*$ in $A(\mathbf{X}^*)$ and $\gamma_1 \in (0, 1)$. It is easy to see that $\hat{G} \in \Pi$.

Notice that $\hat{G}$ is a contraction:

$$\begin{aligned}
\left\|\hat{G}(\mathbf{X}_1, \boldsymbol{\alpha}_1) - \hat{G}(\mathbf{X}_2, \boldsymbol{\alpha}_2)\right\| &= \|[\boldsymbol{\alpha}^* + \gamma_1(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}^*)] - [\boldsymbol{\alpha}^* + \gamma_1(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}^*)]\| = \\
&= \|\gamma_1(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)\| = |\gamma_1| \|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\| \leq \\
&\leq |\gamma_1| \|(\mathbf{X}_1, \boldsymbol{\alpha}_1) - (\mathbf{X}_2, \boldsymbol{\alpha}_2)\|
\end{aligned}$$

Now, consider a contraction of constant $\gamma$, $G(\mathbf{X}_t^e, \boldsymbol{\alpha}_t) : \mathbb{R}^{n \times m} \to \mathbb{R}^m$; thanks to Lemma 13 there is an $\varepsilon$-neighbourhood, $J$, of $G$, (in the $\tau_p$ topology) entirely made up of contractions. We want to show that every such contraction belongs to $\Pi$. For any $G^j \in J$ its restriction $G^j_{\mathbf{X}^*}(\boldsymbol{\alpha}_t)$ is again a contraction of $\mathbb{R}^m$ in itself. As a consequence of Lemma 14 and Remark 2, for each $\eta > 0$ such that

$$\eta(1-\gamma) > \frac{\varepsilon}{1-\varepsilon}$$

we have

$$G^j_{\mathbf{X}^*}(B_\eta(\boldsymbol{\alpha}^*)) \subseteq B_{\frac{\varepsilon}{1-\varepsilon}}(G(B_\eta(\boldsymbol{\alpha}^*))) \subseteq B_{\frac{\varepsilon}{1-\varepsilon}}(B_{\gamma\eta}(\boldsymbol{\alpha}^*)) \subseteq B_\eta(\boldsymbol{\alpha}^*) \qquad (15)$$

which, along with the Contraction Mapping Theorem imply that $G^j_{\mathbf{X}^*}$ has a fixed point, $\boldsymbol{\alpha}^{*j}$, in $\overline{B_\eta(\boldsymbol{\alpha}^*)}$, hence $(\mathbf{X}^*, \boldsymbol{\alpha}^{*j})$ is a fixed point for the system

$$(M^j) : \begin{cases} \mathbf{X}^e_{t+1} = E(\mathbf{X}^e_t, \boldsymbol{\alpha}_t) \\ \boldsymbol{\alpha}_{t+1} = G^j(\mathbf{X}^e_t, \boldsymbol{\alpha}_t) \end{cases}$$

At this point the fact that $G^j$ belongs to $\Pi$, immediately follows the fact that $G^j$ is a contraction. ■

**Remark 3** *The above results are not heavily influenced by the chosen topology. Restricting the analysis to the set (endowed with the $\tau_1$ topology of the $C^1$-uniform convergence) of functions defined on compacts and continuously differentiable, Theorem 15 would still work and the proof could easily be adapted.*

Now to the last step. We now show that every error learning rule with the property (14) can be approximated by a $\mathbf{G} \in \Pi$ with arbitrary precision.

**Theorem 16** *Given the system (7) consider a fixed point $\mathbf{x}^* = F(\mathbf{x}^*)$ and $\mathbf{X}^* = (\mathbf{x}^*, \ldots, \mathbf{x}^*)$. Let $A(\mathbf{X}^*)$ be as in Definition 2. Then, for any $\mathbf{G}_F \in F$ there is a sequence $\{\mathbf{G}_n\}$ ,with $\mathbf{G}_n \in \Pi$ for all $n$, such that*

$$\lim_{n\to+\infty} \mathbf{G}_n(\mathbf{X}^e_t, \boldsymbol{\alpha}_t) = \mathbf{G}_F(\mathbf{X}^e_t, \boldsymbol{\alpha}_t) \qquad \textit{for any } (\mathbf{X}^e_t, \boldsymbol{\alpha}_t)$$

**Proof.**

Consider $\boldsymbol{\alpha}_{t+1} = \mathbf{G}_F(\mathbf{X}^e_t, \boldsymbol{\alpha}_t) \in F$. Let $\boldsymbol{\alpha}^* \in A(\mathbf{X}^*)$. Set

$$\mathbf{G}_n(\mathbf{X}^e_t, \boldsymbol{\alpha}_t) = \frac{1}{n}\boldsymbol{\alpha}^* + \left(1 - \frac{1}{n}\right)\mathbf{G}_F(\mathbf{X}^e_t, \boldsymbol{\alpha}_t)$$

Clearly the sequence converges pointwise to $\mathbf{G}_F$. We show that $\mathbf{G}_n \in \Pi$ for all $n \in \mathbb{N}$:

- First notice that

$$\mathbf{G}_n\left(\mathbf{X}^*,\boldsymbol{\alpha}^*\right)=\frac{1}{n}\boldsymbol{\alpha}^*+\left(1-\frac{1}{n}\right)\mathbf{G}_F\left(\mathbf{X}^*,\boldsymbol{\alpha}^*\right)=\frac{1}{n}\boldsymbol{\alpha}^*+\left(1-\frac{1}{n}\right)\boldsymbol{\alpha}^*=\boldsymbol{\alpha}^*$$

so $(\mathbf{X}^*,\boldsymbol{\alpha}^*)$ is a fixed point for

$$(M_n):\begin{cases}\mathbf{X}^e_{t+1}=\mathbf{E}\left(\mathbf{X}^e_t,\boldsymbol{\alpha}_t\right)\\ \boldsymbol{\alpha}_{t+1}=\mathbf{G}_n\left(\mathbf{X}^e_t,\boldsymbol{\alpha}_t\right)\end{cases}$$

- As for (A2) remark that:

$$\begin{aligned}\left\|\boldsymbol{\alpha}^*-\mathbf{G}_n\left(\mathbf{X}^*,\boldsymbol{\alpha}_t\right)\right\|&=\left\|\boldsymbol{\alpha}^*-\frac{1}{n}\boldsymbol{\alpha}^*-\left(1-\frac{1}{n}\right)\mathbf{G}_F\left(\mathbf{X}^*,\boldsymbol{\alpha}_t\right)\right\|=\\&=\left\|\left(1-\frac{1}{n}\right)\boldsymbol{\alpha}^*-\left(1-\frac{1}{n}\right)\boldsymbol{\alpha}_t\right\|=\left(1-\frac{1}{n}\right)\left\|\boldsymbol{\alpha}^*-\boldsymbol{\alpha}_t\right\|<\\&<\gamma\left\|\boldsymbol{\alpha}^*-\boldsymbol{\alpha}_t\right\|\end{aligned}$$

having set $\left(1-\frac{1}{n}\right)<\gamma<1$.

This completes our proof. ■

## 3.1 Qualifications on topology and compactness.

In this short section we want to discuss how a different assumption about the geometric structure of the set of learning rules of interest could affect the results of the previous section.

Theorem 16 only provides pointwise convergence. In fact, given a function in $\mathbf{G}_F$, a function in $\Pi$ and a point $(\mathbf{X}^*,\boldsymbol{\alpha}_t)$

$$\begin{aligned}&\left\|\mathbf{G}_F\left(\mathbf{X}^*,\boldsymbol{\alpha}_t\right)-\mathbf{G}\left(\mathbf{X}^*,\boldsymbol{\alpha}_t\right)\right\|=\\&=\left\|\boldsymbol{\alpha}_t-\mathbf{G}\left(\mathbf{X}^*,\boldsymbol{\alpha}_t\right)\right\|\geq(1-\gamma)\left\|\boldsymbol{\alpha}^*-\boldsymbol{\alpha}_t\right\|\end{aligned}\tag{16}$$

so that we can make this distance as great as we want if $\|\boldsymbol{\alpha}^*-\boldsymbol{\alpha}_t\|$ (i.e. the domain) is unbounded.

One could still artificially obtain uniform convergence by using, for instance, the following metric. Given a telescopic sequence of compact balls centered in the origin, $K_i$, such that $K_i\subset K_{i+1}$ and $\bigcup_i K_i=\mathbb{R}^m$ define

$$\hat{d}(f,g)=\sum_{i=1}^{\infty}\frac{\sigma_i(f,g)}{2^i}$$

where

$$\sigma_i(f,g) = \frac{d_i(f,g)}{1+d_i(f,g)}$$

and $d_i$ is the metric of $C^1$-uniform convergence on $K_i$. It is clear that this metric cannot be of much practical interest since it annihilates the pointwise distance between functions for points far from the origin: notice that this is precisely the trick to overcome to problem (16). Moreover this *ad-hoc* distance is clearly not independent from the sequence of compacts: different sequences generate different[8] metric spaces.

On the other hand, if we only consider the set of $C^1$ functions, defined on a compact set, endowed with the metric $\tau_1$ of $C^1$-uniform convergence, the theorem 16 can be proved changing the thesis in:

[...

Then, for all $\mathbf{G}_F \in F \subset \Gamma$ there is a sequence $\{\mathbf{G}_n\}$, with $\mathbf{G}_n \in G \subset \Gamma$ for all $n$, such that:

$$\lim_{n\to+\infty} \mathbf{G}_n = \mathbf{G}_F$$

...].

The proof remains unchanged in the existing part. It remains to be proved that the sequence uniformly converges to $\mathbf{G}_F$. Indeed we have:

$$\tau_1(\mathbf{G}_n, \mathbf{G}_F) = \max_{(\mathbf{X}_t^e, \boldsymbol{\alpha}_t)\in K} (\|\mathbf{G}_n - \mathbf{G}_F\| + \|D\mathbf{G}_n - D\mathbf{G}_F\|)$$

where $\|\mathbf{G}(\mathbf{X}_t^e, \boldsymbol{\alpha}_t)\|$ is the euclidean norm in $\mathbb{R}^m$ and $\|D\mathbf{G}(\mathbf{X}_t^e, \boldsymbol{\alpha}_t)\|$ is the norm of the Jacobian matrix $D\mathbf{G}(\mathbf{X}_t^e, \boldsymbol{\alpha}_t)$ defined by

$$\|D\mathbf{G}(\mathbf{X}_t^e, \boldsymbol{\alpha}_t)\| = \max_{\|v\|=1} \|D\mathbf{G}(\mathbf{X}_t^e, \boldsymbol{\alpha}_t)\, v\|$$

But

$$\|\mathbf{G}_n - \mathbf{G}_F\| = \frac{1}{n} \|\boldsymbol{\alpha}^* - \mathbf{G}_F(\mathbf{X}_t^e, \boldsymbol{\alpha}_t)\|$$

and

$$\|D\mathbf{G}_n - D\mathbf{G}_F\| = \|D(\mathbf{G}_n - \mathbf{G}_F)\| = \frac{1}{n} \|D(\boldsymbol{\alpha}^* - \mathbf{G}_F(\mathbf{X}_t^e, \boldsymbol{\alpha}_t))\|$$

[8] These different metric spaces, though, are topologically equivalent; they also generate on each compact set a topology which is equivalent to that of $C^1$-uniform convergence.

hence

$$\tau_1(\mathbf{G}_n, \mathbf{G}_F) = \frac{1}{n} \max_{(\mathbf{X}_t^e, \boldsymbol{\alpha}_t) \in K} \left( \|\boldsymbol{\alpha}^* - \mathbf{G}_F(\mathbf{X}_t^e, \boldsymbol{\alpha}_t)\| + \|D(\boldsymbol{\alpha}^* - \mathbf{G}_F(\mathbf{X}_t^e, \boldsymbol{\alpha}_t))\| \right)$$

and, because both norms are bounded, for any $\varepsilon > 0$ there is $n_\varepsilon$ such that for all $n > n_\varepsilon$ we have

$$\tau_1(\mathbf{G}_n, \mathbf{G}_F) < \varepsilon$$

which is the claimed result.

# 4 Examples.

In this section we provide examples that should help understanding the possible practical relevance of the theoretic results of the previous sections. The examples we show here are drawn from the economic literature and are in some sense typical.

We begin considering a very simple case of adaptive expectations. In an economy with infinite past, let's consider expectations as formed through a weighted mean of observed data

$$x_{t+1}^e = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i x_{t-i} \qquad \alpha \in (0,1)$$

which corresponds to

$$x_{t+1}^e = x_t^e + \alpha(x_t - x_t^e)$$

Suppose now that the parameter $\alpha$ is endogenously determined as follows

$$\begin{cases} x_{t+1}^e = x_t^e + \alpha_t(x_t - x_t^e) \\ \alpha_{t+1} = g(x_t - x_t^e, \alpha_t) \end{cases} \tag{17}$$

Endogenizing the parameter as in (17) can be interesting for many reasons. The idea is that the agents, although boundedly rational, might recognize the opportunity of tuning the weight of observed data as more evidence becomes available: such process should be captured by the law $g(.)$ that appears in (17). To show a simple case in which this happens consider the same model, but with a finite past. Then we can rewrite the correspondent expectation mechanism as

$$x_{t+1}^e = \frac{1-\rho}{1-\rho^{T+1}} \sum_{i=0}^{T} \rho^i x_{t-i} \qquad \rho \in (0,1) \tag{18}$$

which leads to

$$\left\{ \begin{array}{l} x^e_{t+1} = x^e_t + \alpha_t \left( x_t - x^e_t \right) \\ \alpha_{t+1} = \frac{\alpha_t}{\rho + \alpha_t} \qquad \alpha_0 = 1 \end{array} \right. \tag{19}$$

This kind of expectations formation mechanism is known as *exponential smoothing* or *fading memory learning*. Clearly the behavior of a system like (19) depends strongly on the way the state variable is linked to the expectations. In the literature one can find several examples of this type; see for instance Bischi-Gardini [4], Bischi-Naimzada [5], Barucci [3], Pötzelberger and Sögner [30]. Remark that the fading memory learning process (19) satisfies Assumptions (A1) and (A2) of Theorem 6. Therefore in general, under suitable properties for the temporary equilibrium map, application of Theorem 9 shows that a fading memory learning process enhances stability (according to Definition 3) with respect to ordinary adaptive expectations.

Another simple example can be obtained from the case of fading memory when the parameter $\rho$ tends to one:

$$x^e_{t+1} = \frac{1}{t+1} \sum_{i=0}^{t} x_i \tag{20}$$

In this case expectations are formulated as a simple arithmetic mean of the last $t$ observations. This type of learning process has been used in Bray [7] and in Lucas [24]. Notice that (20) can also be written as

$$\left\{ \begin{array}{l} x^e_{t+1} = x^e_t + \alpha_t \left( x_t - x^e_t \right) \\ \alpha_{t+1} = \frac{\alpha_t}{1 + \alpha_t} \qquad \alpha_0 = 1 \end{array} \right. \tag{21}$$

The system (21) satisfies (A1) of Theorem 6 but not (A2), in the sense that one cannot find a $\gamma \in (0, 1)$ which works uniformly for all times ($\gamma = 1$ would actually work but it is not enough). The results of the previous sections can still be applied to the system (21) if we slightly weaken Assumption (A2) (and require that $A(x^*)$ be non-empty) [9].

[9] This can be done for instance in either of the two following ways:

(A2*bis*) For any $\delta > 0$ (and a neighborhood of $\alpha^*$ of radius $\delta$, $I_\delta$) there is $\gamma \in (0, 1)$ such that

$$\|\alpha^* - g(0, \alpha_t)\| \leq \gamma \|\alpha^* - \alpha_t\| \qquad \text{whenever } \alpha_t \notin I_\delta$$

(A2*ter*) There is a continuous function $P : \mathbb{R} \setminus \{\alpha^*\} \to (0, 1)$, $\alpha \longmapsto P(\alpha)$ such that

$$\|\alpha^* - g(0, \alpha_t)\| \leq P(\alpha_t) \|\alpha^* - \alpha_t\| \qquad \text{whenever } \alpha_t \neq \alpha^*$$

## 4.1 A generalization of "fading memory".

Going back to the fading memory learning rules of (18) and (19), we can develop an interesting generalization to account for the (possible) capability of agents to perform some sort of "endogenous" assessment of the available data. Recall that in this setting agents form their forecasts as a weighted average of old observed data and the parameter $\rho$ $(0 \leq \rho \leq 1)$ can be thought of as a measure of the significance of the time series. It is reasonable to think that the attitude of the subject may vary over time, reflecting her confidence in the reliability of the available observations as instruments to provide accurate predictions. Therefore a "low" value for $\rho$ (i.e. $\rho$ near to 0) should be associated to moments of lack of confidence in the old data and vice versa. In particular we will make the following behavioral assumptions: 1) At each time, traders decide the relative weight of older data with respect to the last $k$ observations and forecasts (for example considering last $k$ forecast errors); 2) they remember their past negative evaluations. A precise formalization of this is as follows. Expectations are given by:

$$x_{t+1}^e = \frac{x_t + \rho_t x_{t-1} + \rho_t \rho_{t-1} x_{t-2} + \cdots + \left(\rho_t \rho_{t-1} \cdots \rho_1\right) x_0}{1 + \rho_t + \rho_t \rho_{t-1} + \cdots + \rho_t \rho_{t-1} \cdots \rho_1} \tag{22}$$

where

$$\rho_t = H\left(x_t, \ldots, x_{t-k+1}, x_t^e, \ldots, x_{t-k+1}^e\right) = H\left(X_t, X_t^e\right)$$

and, in a more compact form, the system is

$$\begin{cases} x_{t+1}^e = \dfrac{x_t + \sum_{j=0}^{t-1} \prod_{i=0}^{j} \rho_{t-i} x_{t-i-j}}{1 + \sum_{j=0}^{t-1} \prod_{i=0}^{j} \rho_{t-i}} \\ \rho_{t+1} = H\left(X_{t+1}, X_{t+1}^e\right) \end{cases}$$

which corresponds to

$$\begin{cases} x_{t+1}^e = x_t^e + \alpha_t \left(x_t - x_t^e\right) \\ \alpha_{t+1} = \frac{\alpha_t}{\rho_t + \alpha_t} \qquad \alpha_0 = 1 \\ \rho_{t+1} = H\left(X_{t+1}, X_{t+1}^e\right) \end{cases} \tag{23}$$

In either case the results of the previous Sections, and in particular 6, still hold under the milder assumptions (A2*bis*) or (A2*ter*). Notice that a specification of the learning rule as in (21), i.e. $\alpha_{t+1} = \frac{\alpha_t}{1+\alpha_t}$ satisfies both (A2*bis*) and (A2*ter*): for instance one can set $P\left(\alpha_t\right) = \left\|1 + \alpha_t\right\|^{-1}$ to satisfy (A2*ter*).

Notice that the function $H(.)$ embeds all the information about agents psychology and therefore its shape should strongly depend on the economic phenomenon under study. Under rather mild assumptions on $H(.)$ though, we can resort to the results of the previous sections.

**Proposition 17** *Given the system (23), let* $\rho^* = H(X^*, X^*)$ *and* $\alpha^* = \frac{1}{\sum_{k=0}^{+\infty}(\rho^*)^k}$. *Assume that* $H(X_t, X_t^e) \leq \theta < 1$ *for all* $X_t, X_t^e$. *Then the system (23) satisfies the assumption (A2).*

**Proof.** In the Appendix.

To give a more precise idea of where this could lead to, we will need to impose stronger conditions on the shape of the function $H(.)$ i.e. to make a further behavioural assumption; in Colucci-Valori [11] we suppose that agents react to the last error, returning a "small" $\rho$ for large errors and vice-versa. Marimon and Sunder [29] provides an experimental-based support for this assumption. As an example we might consider the following:

$$H(x_t, x_t^e) = ke^{-[h(x_t - x_t^e)]^2} + d$$

with $h, d, k > 0$ and $k + d < 1$. Notice that it is also possible to determine conditions which secure that $\alpha^* \in A(x^*)$ so that Theorem 6 applies.

We carry on a detailed treatment of the type of analysis sketched here in Colucci-Valori [11] where we also develop numerical evidence for a Cobweb-style market which is an extension of the model studied in Hommes [21] and [22].

## 4.2 OLS learning

We now turn to the very well known case of recursive least square learning. Some recent references include Bullard [9], Grandmont [20], Schönhofer [32, 33], Tuinstra [35], Evans and Honkapohja [13] and Chatterji-Chattopadhyay [10].

Suppose that the *actual law of motion* is given by

$$x_s = F(x_t^e, \ldots, x_{t-P}^e) \tag{24}$$

while the traders, in order to form their expectations, refer to a *perceived law of motion* of the type

$$x_{t+1} = \beta_t x_t \tag{25}$$

where the coefficient $\beta$, representing the expected growth rate of the state variable, is estimated through ordinary least squares (using all previous observations)

$$\beta_t = \frac{\sum_{i=1}^{s} x_{i-1}x_i}{\sum_{i=1}^{s} x_{i-1}^2}$$

It can be shown (e.g. Ljung and Soderstrom [25] and Marcet and Sargent [27]) that the following recursive formulation is also possible:

$$\begin{cases} \beta_{t+1} = \beta_t + \alpha_t \left( \frac{x_{s+1}}{x_s} - \beta_t \right) \\ \alpha_{t+1} = \frac{\alpha_t}{\frac{x_s^2}{x_{s+1}^2} + \alpha_t} \end{cases} \tag{26}$$

Remark that a fixed point for (26) is either of the form $\left(\beta^*, 1 - (\beta^*)^{-2}\right)$ or of the form $(\beta, 0)$.

The first steady state requires that $F(x_t^e, \dots, x_{t-P}^e) = x_s = \beta^* x_{s-1}$ so that perceived and actual law of motion coincide. This coincidence does not depend on the specification of the temporary equilibrium map. Now set $Y_t = \frac{x_{s+1}}{x_s}$. Assumption (A2) is well defined provided that $X^*$ of Definition 2 is substituted with the equilibrium value of $Y_t$ which correspond to $Y^* = \beta^*$. With this unimportant modification and restricting our analysis to the case $\beta > 1$, we obtain, as a corollary of Proposition 17, that assumption (A2) is satisfied by system (26).

In the second case, (A2) does not hold. Given a particular temporary equilibrium map we could try to prove a weaker version of it, as in the case of the Bray learning rule; anyway nothing can be said in general.

Suppose now that the temporary equilibrium map is such that $y_t = \left(\frac{x_{s+1}}{x_s}\right) = F\left(\left(\frac{x_t}{x_{t-1}}\right)^e\right)$. If this is the case we consider the growth rate of

$x_t$ as the state variable and the system reduces to[10]

$$\begin{cases} \beta_{t+1} = \beta_t + \alpha_t \left( y_t - \beta_t \right) \\ \alpha_{t+1} = \frac{\alpha_t}{y_t^{-2} + \alpha_t} \end{cases} \tag{27}$$

As $\beta$ represents the expected value of $y$ (remember that agents are supposed to believe in (25) which is also used in order to perform the forecasts), we can immediately check that the system (27) is a particular case of (23) with $H(\cdot, \cdot) = x_t^{-2}$.

# 5 Appendix.

**Proof.** of Proposition 12.

Let $B_j$ and $K_i$ be two telescopic sequences of compact sets such that $\bigcup_j B_j = \bigcup_i K_i = \mathbb{R}^m$. Let $d_B$ and $d_K$ be the distances defined through $\{B_j\}$ and $\{K_i\}$ respectively. Suppose that $f$ and $g$ be such that $d_k(f, g) \neq d_B(f, g)$, and without loss of generality $d_k(f, g) < d_B(f, g)$; then for any $\varepsilon > 0$ there is $n(\varepsilon)$ such that

$$\left| \frac{\max_{x \in B_j} \left( \| f(x) - g(x) \| + \| Df(x) - Dg(x) \| \right)}{1 + \max_{x \in Bj} \left( \| f(x) - g(x) \| + \| Df(x) - Dg(x) \| \right)} - d_B(f, g) \right| < \varepsilon \tag{28}$$

[10] An example of this case, obviously, is given by the overlapping generations model. In a pure consumption framework, with two-period-living agents and constant money growth the temporary equilibrium map is given by:

$$\frac{p_t}{p_{t-1}} = \pi_t = \theta \frac{S\left(\pi_{t-1}^e\right)}{S\left(\pi_t^e\right)}$$

where $p$ is the price level, $\pi$ is the inflation rate, $\theta$ is the money growth rate and $S$ the saving function.

If agents have a perceived law of motion of the kind of (25) and estimate the coefficient $\beta$ through ordinary least squares we obtain:

$$\begin{cases} \beta_{t+1} = \beta_t + \alpha_t \left( \theta \frac{S(\beta_{t-1})}{S(\beta_t)} - \beta_t \right) \\ \alpha_{t+1} = \frac{\alpha_t}{\left( \theta \frac{S(\beta_{t-1})}{S(\beta_t)} \right)^{-2} + \alpha_t} \end{cases}$$

where $\beta_t$ represents the expected (OLS-estimated) rate of inflation.

For a more comprehensive analysis of this model see Bullard [9] Schönofer [32] and Tuinstra [35].

for all $j > n(\varepsilon)$. Hence, there is $\bar{x}$ such that

$$\left| \frac{(\|f(\bar{x}) - g(\bar{x})\| + \|Df(\bar{x}) - Dg(\bar{x})\|)}{1 + (\|f(\bar{x}) - g(\bar{x})\| + \|Df(\bar{x}) - Dg(\bar{x})\|)} - d_B(f, g) \right| < \varepsilon$$

For such a $\bar{x}$ there is a $K_{i_0}$ such that $\bar{x} \in K_{i_0}$, moreover

$$\frac{\max_{x \in K_{i_0}} (\|f(x) - g(x)\| + \|Df(x) - Dg(x)\|)}{1 + \max_{x \in K_{i_0}} (\|f(x) - g(x)\| + \|Df(x) - Dg(x)\|)} < d_K(f, g) \qquad (29)$$

So, putting together (28) and (29), we have

$$d_K(f, g) > d_B(f, g) - \varepsilon$$

which contradict the assumption, as $\varepsilon$ is arbitrarily small. ■

**Lemma 18** *Let $f \in \mathcal{F}$. The following statements are equivalent:*
*a) $f$ is a contraction of constant $\gamma$, i.e.*

$$d(f(x), f(y)) \leq \gamma d(x, y)$$

*($d$ is the euclidean distance)*
*b) $f$ is such that*

$$\|Df(x)\| < \gamma \qquad \text{for any } x \in \mathbb{R}^n$$

**Proof.** (a $\Rightarrow$ b) :

Let $f$ a contraction of constant $\gamma$, we must show that $\|Df(x)v\| < \gamma$ for all $\|v\| = 1$. The inequality

$$\|Df(x)v\| = \left\| \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} \right\| < \lim_{t \to 0} \frac{\gamma \|tv\|}{|t|} = \gamma$$

proves this.

(b $\Rightarrow$ a) :

Let us estimate $\|f(x) - f(y)\|$. We have:

$$f(y) - f(x) = \int_0^1 \frac{d}{dt} f(x + t(y - x))\, dt = \int_0^1 Df(x + t(y - x))(y - x)\, dt$$

hence

$$\|f(x) - f(y)\| \leq \int_0^1 \|Df(x + t(y - x))\| \|y - x\|\, dt < \int_0^1 \gamma \|y - x\|\, dt = \gamma \|y - x\|$$

■

**Proof.** of Lemma 13

Let $f$ be a contraction of constant $\gamma$ and $g \in \mathcal{F}$ such that $\tau_p(f,g) < \varepsilon$. From the definition of $\tau_p$ it follows that for any $x \in \mathbb{R}^n$ (see Remark 2)

$$\|Df(x) - Dg(x)\| < \frac{\varepsilon}{1-\varepsilon} < 2\varepsilon \qquad \text{if} \qquad \varepsilon < \frac{1}{2}$$

Therefore for $\varepsilon < \frac{1}{2}$

$$|\|Df(x)\| - \|Dg(x)\|| < \|Df(x) - Dg(x)\| < 2\varepsilon$$

so

$$\|Dg(x)\| < \|Df(x)\| + 2\varepsilon$$

By Lemma 18, given $\varepsilon$ such that $\gamma + 2\varepsilon < 1$, it follows that $g$ is a contraction, as required. ■

**Proof.** of Proposition 17.

To see that (A2) is satisfied observe that $\alpha^* = \frac{1}{\sum_{k=0}^{+\infty}(\rho^*)^k} = 1 - \rho^*$; therefore

$$\begin{aligned}
|\alpha^* - G(x^*, \alpha_t)| &= \left|\alpha^* - \frac{\alpha_t}{H(x^*, x^*) + \alpha_t}\right| = \left|\alpha^* - \frac{\alpha_t}{\rho^* + \alpha_t}\right| = \\
&= \left|\alpha^* - \frac{\alpha_t}{1 - \alpha^* + \alpha_t}\right| = \left|\frac{\alpha^* - {\alpha^*}^2 + \alpha^*\alpha_t - \alpha_t}{1 - \alpha^* + \alpha_t}\right| = \\
&= \left|\frac{\alpha^*(1-\alpha^*) - \alpha_t(1-\alpha^*)}{1 - \alpha^* + \alpha_t}\right| = \left|\frac{(1-\alpha^*)(\alpha^* - \alpha_t)}{1 - \alpha^* + \alpha_t}\right| = \\
&= \frac{1-\alpha^*}{1 - \alpha^* + \alpha_t}|\alpha^* - \alpha_t|
\end{aligned}$$

We now show that there is $\gamma$ such that $0 < \frac{1-\alpha^*}{\alpha_t + 1 - \alpha^*} < \gamma < 1$ for all $t$, or equivalently, that there is $\eta > 0$ such that $\alpha_t \geq \eta$ for all $t$. In the second equation of (23), substituting recursively, we get

$$\alpha_t = \frac{1}{\frac{\prod_{i=0}^{t-1}\rho_i}{\alpha_0} + 1 + \sum_{j=0}^{t-2}\prod_{i=0}^{j}\rho_{t-1-i}} \geq \frac{1}{\frac{\theta^t}{\alpha_0} + \sum_{i=0}^{t-1}\theta^i} \geq \frac{1}{\frac{1}{\alpha_0} + \frac{1}{1-\theta}} \overset{def}{=} \eta$$

where the first inequality stems from the fact that $H$ is non-negative and upper bounded by $\theta$. We can then choose $\gamma = \frac{1-\alpha^*}{\eta + 1 - \alpha^*} < 1$, for which (A2) is satisfied for all $\alpha_t$. ■

## References


[1] W.B. Arthur, J.H. Holland, B. LeBaron, R. Palmer, and P. Tayler: Asset Pricing Under Endogenous Expectations in an Artificial Stock Market. In W.B. Arthur, S.N. Durlauf, and D. Lane Eds., *The Economy as an Evolving Complex Systems II*, Addison-Wesley Pub., Redwood City, (1997).

[2] Y. Balasko and D. Royer: Stability of competitive equilibrium with respect to recursive and learning processes. *Journal of Economic Theory*, **68**, 319-348 (1996).

[3] E. Barucci: Exponentially fading memory learning in forward-looking economic models. *Journal Of Economic Dynamics And Control* **(24)5-7,** 1027-1046 (2000).

[4] G.I. Bischi, L. Gardini: Mann Iterations reducible to plane endomorphisms. *Quaderni di Economia Matematica e Statistica*, **36**, Università di Urbino (1995).

[5] G.I. Bischi, A. Naimzada: Global analysis of a non-linear model with learning. *Economic Notes* **26(3)**, 445-476 (1997)

[6] M. Bray: Learning, estimation and the stability of rational expectations. *Journal of Economic Theory*, **26**, 318-339 (1979).

[7] M. Bray: Convergence to rational expectations equilibrium. In R. Friedman and E.S. Phelps (eds.), *Individual forecasting and aggregate outcomes*, Cambridge University Press, (1983).

[8] W.A. Brock, C.H. Hommes: A rational route to randomness. *Econometrica*, **65**(5), 1059-1095 (1997).

[9] J. Bullard: Learning equilibria. *Journal of Economic Theory* **64**, 468-485 (1994).

[10] S. Chatterji, S. Chattopadhyay: Global stability in spite of “local instability” with learning. *Journal of Mathematical Economics*, **33**, 155-165 (2000)

[11] D. Colucci and V. Valori: A generalization of “fading memory”, with an application to the Cobweb model. In preparation, (2001).

[12] R. Devaney: *An introduction to chaotic dynamical systems*, Addison Wesley, (1987).

[13] G.W. Evans, S. Honkapohja: *Learning and expectations in macroeconomics*. Princeton University Press (2001).

[14] G.W. Evans, S. Honkapohja, R. Marimon: Convergence of monetary inflation models with heterogeneous learning rules. Forthcoming on *Macroeconomic Dynamics*. (2000).

[15] G.W. Evans, G. Ramey: Expectation, calculation and macroeconomic dynamics. *American Economic Review*, **82**, 207-224 (1992).

[16] G.W. Evans, G. Ramey: Calculation, adaptation and rational expectations. *Macroeconomic Dynamics*. **2**, 156-182 (1998).

[17] G. Fuchs: Is error learning behaviour stabilizing? *Journal of Economic Theory*, **20**, 300-317 (1979).

[18] J. M. Grandmont: Temporary general equilibrium theory. *Econometrica*, **45**, 535-572 (1977).

[19] J. M. Grandmont: On endogenous competitive business cycles. *Econometrica*, **53**, 995-1045 (1985).

[20] J. M. Grandmont: Expectations formation and the stability of large socioeconomic systems. *Econometrica*, **66**, 741-781 (1998).

[21] C. H. Hommes: Adaptive learning and road to chaos. *Economics Letters*, **36**, 127-132 (1991).

[22] C. H. Hommes: Dynamics of the cobweb model with adaptive expectations and nonlinear supply and demand. *Journal of Economic Behavior and Organization* **24**, 315-335 (1994).

[23] C. H. Hommes: On the consistency of backward looking expectations. The case of the cobweb. *Journal of Economic Behavior and Organization* **33**, 333-362 (1998).

[24] R. Lucas: Adaptive behaviour and economic theory. *Journal of Business* **59(4)**, 401-425 (1986).

[25] L. Ljung and T. Soderstrom: *Theory and practice of recursive identification.* MIT Press, (1983)

[26] W. R. Mann: Mean value methods in iteration. *Proceedings of the A.M.S.* **4**, 506-510 (1953).

[27] A. Marcet and T. Sargent: Convergence of least square learning in self-referential stochastic models. *Journal of Economic Theory* **48**, 337-368 (1989).

[28] R. Marimon: Learning from learning in economics. In D. Kreps, K. Wallis, Editors. *Advances in economics and econometrics: theory and applications.* Vol. 1. Cambridge University Press (1997).

[29] R. Marimon, S. Sunder: Expectations and learning under alternative monetary regimes: an experimental approach. *Economic Theory* **4(1)**, 131-62 (1994)

[30] K. Pötzelberger and L. Sögner: Stochastic equilibrium: learning by exponential smoothing. *SFB Working Paper*, Vienna University of Economics and Business Administration, n. 67 (2000).

[31] T.J. Sargent: *Bounded rationality in macroeconomics.* Oxford University Press (1993).

[32] M. Schönhofer: Chaotic learning equilibria. *Journal of Economic Theory*, **89**, 1-20 (1999).

[33] M. Schönhofer: Can agents learn their way out of chaos? *Journal of Economic Behavior and Organization*, **44**, 71-83 (2001).

[34] R. Slonim: Learning rules of thumb or learning more rational rules. *Journal of Economic Behavior and Organization*, **38**, 217-236 (1999).

[35] J. Tuinstra: Learning in overlapping generations models. *CeNDEF Working Paper* **99-03**, University of Amsterdam (1999).

[36] M. Woodford: Nonstandard Indicators for Monetary Policy: Can Their Usefulness be Judged from Forecasting Regressions? *Monetary Policy*, Mankiw ed., University Chicago Press (1994).