Ferenc Szidarovszky
Gian Italo Bischi *Editors*

Games and Dynamics in Economics

Essays in Honor of Akio Matsumoto

Springer

Path Dependence in Models with Fading Memory or Adaptive Learning

Gian Italo Bischi, Laura Gardini and Ahmad Naimzada

Abstract We consider a learning mechanism where expected values of an economic variable in discrete time are computed in the form of a weighted average that exponentially discounts older data. Also adaptive expectations can be expressed as weighted sums of infinitely many past states, with exponentially decreasing weights, but these are not averages since the weights do not sum up to one for any given initial time. These two different kinds of learning, which are often considered as equivalent in the literature, are compared in this paper. The statistical learning dynamics with exponentially decreasing weights can be reduced to the study of a two-dimensional autonomous dynamical system, whose limiting sets are the same as those obtained with adaptive expectations. However, starting from a given initial condition, different transient dynamics are obtained, and consequently convergence to different attracting sets may occur. In other words, even if the two different kinds of learning dynamics have the same attracting sets, they may have different basins of attraction. This implies that local stability results are not sufficient to select the kind of long-run dynamics since this may crucially depend on the initial conditions. We show that the two-dimensional discrete dynamical system equivalent to the statistical learning with fading memory is represented by a triangular map with denominator which vanishes along a line, and this gives rise to particular structures of their basins of attraction, whose study requires a global analysis of the map. We discuss some examples motivated by the economic literature.

Keywords Economic dynamics · Learning · Expectations · Attractors · Basins · Bifurcations

G. I. Bischi (✉) · L. Gardini
Dipartimento di Economia, Società, Politica (DESP), Università di Urbino, Urbino, Italy
e-mail: gian.bischi@uniurb.it

L. Gardini
e-mail: laura.gardini@uniurb.it

A. Naimzada
Dipartimento di Economia, Management e Statistica, Università Bicocca, Milano, Italy
e-mail: ahmad.naimzada@unimib.it

F. Szidarovszky and G. I. Bischi (eds.), *Games and Dynamics in Economics*,
https://doi.org/10.1007/978-981-15-3623-6_3

1 Introduction

Many dynamic models involve memory of past states to determine the future time evolution of systems in physics, engineering, natural sciences and economics. The inclusion of past history in the time evolution adds nontrivial complexities, balancing the advantage of dealing with more realistic models. In economics the inclusion of memory in modeling human decisions may be considered as a method to represent learning processes (see e.g. Hommes et al. 2012). The effects of memory in continuous time models of oligopoly markets has recently been analyzed by Matsumoto and Szidarovszky in a series of papers dealing with problems of stability of equilibrium points as time lags are varied, as well as bifurcations leading to dynamic complexities whose consequences are studied both by analytical and numerical methods, see e.g. Matsumoto and Szidarovszky (2018), and Matsumoto and Szidarovszky (2015), as well as Matsumoto (2017). Dynamic models involving delays often generate dynamical systems of infinite dimension. However, some particular kinds of distributed delays have been introduced, expressed by integral terms with kernels (denoted as gamma functions) characterized by an exponential decay going back in the past, that allow to transform an integrodifferential equation into an expanded set of ordinary differential equations of finite dimension (see e.g. Cushing 1978; MacDonald 1978; Chiarella 1991).

In this paper we consider discrete time dynamic models, often used to describe social and economic systems characterized by event-driven time, simulate to describe agents that take decisions by considering past information with exponentially distributed weights, i.e. an exponentially fading memory. In particular, we consider economic models that involve agents' expectations about the future states of the system, and are formulated as mappings from beliefs to realizations, such as $x_t = F\left(x_t^{(e)}\right)$ or $x_t = F\left(x_{t+1}^{(e)}\right)$, where $x_t^{(e)}$ and $x_{t+1}^{(e)}$ represent agents' expectations about current or future states respectively. In order to close the model one must introduce a learning mechanism by which agents make forecastings on the basis of the past history of the system. In this paper we consider one-dimensional models with expectations endowed with two kinds of learning: The first is known as *adaptive learning* (see e.g. Hommes 2013 and references therein) where expectations are obtained by assuming that at each time the expected value is a weighted average of the previous forecast and the previous observed value; the second, obtained by assuming that at each time period the agents compute the expected value as an average of the past realized values, starting from a given initial time $t = 0$, is sometimes called *statistical learning* (see e.g. Guesnerie and Woodford 1992).

Both learning mechanisms share the same equilibrium points of the corresponding model with rational expectations (or perfect foresight) $x_t^{(e)} = x_t$ for each t, that is, assuming that agents are able to anticipate the future outcomes, so that expectations are fulfilled at each time. So, it is interesting to consider the problem of stability of such "rational equilibria" under these learning mechanisms. The *Rational Expectations* (RE) hypothesis, based on the assumptions that agents have

complete knowledge of the economic model and fully exploit all the available pieces of information, has been criticized from many points of view, mainly because the assumptions behind the RE paradigm seem to require too much agents' rationality. So, models with boundedly rational agents that converge in the long run to a rational equilibrium may be seen as an evolutionary interpretation of rationality, and some authors say that in this case the boundedly rational agents are able to learn, in the long run, what rational agents already know under very pretentious rationality assumptions (see e.g. Fudenberg and Levine 1998). However, it may happen that under different starting conditions (or as a consequence of exogenous perturbations) the same adaptive process leads to non-rational equilibria as well, i.e. equilibrium situations which are different from the ones forecasted under the assumption of full rationality, as well as to dynamic attractors characterized by endless asymptotic fluctuations or unfeasible evolutions. The coexistence of several attracting sets, each with its own basin of attraction, gives rise to path dependence, irreversibility, hysteresis and other nonlinear and complex phenomena commonly observed in real systems as well as in laboratory experiments. So, stability arguments under some learning dynamics are often used as *equilibrium selection* criteria.

In this paper we consider a particular statistical learning in which the agents discount older data by making weighted averages with exponentially decreasing weights (see Bischi and Gardini 1996; Bischi and Naimzada 1997), so it is the analogous of an exponentially decreasing gamma kernel often used in continuous time dynamic models with distributed delays. Moreover, the discrete fading memory analyzed in this paper includes, as a limiting case, the learning process proposed by Bray (1983). Even adaptive expectations can be expressed as weighted sums of infinitely many past states, with exponentially decreasing weights, but these are not averages since the weights do not sum up to one for any finite initial time. These two different kinds of learning are often considered as equivalent in the literature, because they assume the same form as $t \to +\infty$. Indeed, statistical learning dynamics with exponentially decreasing weights can be reduced to the study of a two-dimensional autonomous dynamical system, whose limiting sets are the same as those obtained with adaptive expectations. However, starting from a given initial condition, different transient dynamics are obtained, and consequently convergence to different attracting sets may occur. In other words, even if the two different kinds of learning dynamics have the same attracting sets, they may have different basins of attraction. In situations of *multistability*, i.e. when several coexisting attractors are present, local stability results are not sufficient to provide selection criteria since this may crucially depend on the initial conditions. Hence, adaptive and statistical learning may give different results when the problem of equilibrium selection arises. Moreover, we show that the two-dimensional discrete dynamical system equivalent to the statistical learning with fading memory is represented by an iterated two-dimensional triangular map with denominator which vanishes along a line, and this gives rise to particular structures of their basins of attraction, whose study requires a global analysis of the map following a stream of literature dealing with maps which are not defined in the whole phase space due to the presence of vanishing denominators, see Bischi et al. (1999), Bischi et al. (2003), and Bischi et al. (2005). In particular, we show that the structure of the

basins is strongly influenced by the presence of particular points, called *focal points* in Bischi and Gardini (1997) and Bischi et al. (1999), whose existence, in the case of models with expectations, is related to the presence of fixed points of the map F, which are rational expectations equilibria. These global properties are specific to discrete time models, in the sense that they cannot be observed in continuous time models with delays.

The plan of the paper is the following. In Sect. 2 we compare the mathematical form of models with adaptive expectations and those with statistical learning. In Sect. 3 their dynamical properties are studied, in particular those of statistical learning with fading memory analyzed through the study of an equivalent two-dimensional iterated map, with particular emphasis on the study of the basins of attraction and their global bifurcations specific to maps with a vanishing denominator. In Sect. 4 some examples are discussed, Sect. 5 concludes and suggests some possible extensions.

2 From Beliefs to Realizations: Rational Expectations and Learning Dynamics

Let us consider one-dimensional discrete time economic models represented by mappings from expected values to realized values of the same period, i.e. the outcome of the state variable x_t at time t is a function of the value $x_t^{(e)}$ which agents expect, at the same time t, for the state variable, computed by the agents on the basis of the information held at the previous time $(t-1)$

$$x_t = F\left(x_t^{(e)}\right) \tag{1}$$

If the agents have *Rational Expectations* (RE), which in a deterministic framework means that they are endowed by *Perfect Foresight* (PF), the expected values coincide with the realized values at each time

$$x_t^{(e)} = x_t \qquad \forall t \tag{2}$$

If (2) is inserted into (1) we get

$$x_t = F\left(x_t\right)$$

which means that only a *Rational Expectations Equilibrium* (REE) is a fixed point of the map F. It is often argued that the assumption of rational expectations is too strong, since economic models should take into account human limited ability to make forecastings. This leads to the weaker assumption of *Bounded Rationality* (BR) which assumes that agents compute the expected values $x_t^{(e)}$ by some *learning mechanism* based on past experience, i.e.

$$x_t^{(e)} = \Psi\left(x_{t-1}, x_{t-2}, ..., x_{t-1}^{(e)}, x_{t-2}^{(e)}, ...\right) \tag{3}$$

This assumption is not only introduced by claiming that BR is more realistic than RE, but it is often used as a *REE justification* or as a dynamic mechanism for *equilibrium selection* when several RE equilibria exist (see e.g. Marimon 1997). Of course, this requires that a REE must also be an equilibrium for the model with learning. In this context, the local stability of a REE with respect to the dynamics induced by bounded rationality learning is commonly referred to as an *evolutive* explanation of the RE solution, see Guesnerie and Woodford (1992). Moreover, some REEs may be more likely to be reached than other ones when some learning mechanism is introduced (some may be not reached at all if they are unstable under the chosen learning process). Even more interesting situations of equilibrium selection arise when there are attractors of the dynamics with learning which do not exist with RE. This leads, for the dynamics with bounded rationality, to situations of coexistence of attractors which are rational, i.e. also exist for the model with RE assumption, with attractors which are non rational, i.e. asymptotic evolutions which do not exist under the assumption of RE, so that the long-run behavior is characterized by agents which continue to make wrong forecastings.

The selection of the attractor, in particular the convergence to a rational or a non rational attractor, may depend on the initial condition, i.e. from the boundaries among the different basins of attraction. This aspect has been rather neglected in the literature because it requires a global analysis of the dynamics with learning.

2.1 Adaptive Learning and Reduction to One-Dimensional Dynamics

A simple and frequently used learning mechanism is given by the *adaptive expectations*, expressed by

$$x_{t+1}^{(e)} = x_t^{(e)} + \alpha \left(x_t - x_t^{(e)} \right) \quad 0 \leq \alpha \leq 1. \tag{4}$$

i.e. for each time $t = 0, 1, \ldots$. the value $x_{t+1}^{(e)}$ expected for the next period $(t + 1)$ is obtained by "adapting" the previous forecasting $x_t^{(e)}$ in the direction of the corresponding observed value x_t, with a speed of adjustment α. Rearranging (4) the new expected value $x_{t+1}^{(e)}$ can be expressed as a convex combination (i.e. a weighted average) of the previous expected value $x_t^{(e)}$ and the currently observed value x_t

$$x_{t+1}^{(e)} = (1 - \alpha)\, x_t^{(e)} + \alpha x_t \quad 0 < \alpha < 1 \tag{5}$$

We can observe that the limiting case $\alpha = 1$ corresponds to static (or naive) expectations

$$x_{t+1}^{(e)} = x_t \tag{6}$$

and decreasing values of α correspond to higher inertia in updating the previously expected value according to the more recent observation.

Using (5) repeatedly, the adaptively expected value can be expressed as

$$x_{t+1}^{(e)} = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k \, x_{t-k} \tag{7}$$

i.e. infinitely many past realizations are considered, with weights exponentially decreasing as more remote past values are considered (decreasing as the terms of a geometric sequence of ratio $(1-\alpha)$). Some authors call (7) a weighted average of the values observed in the past, but it is important to remark that (7) *cannot be considered as an average*, since the weights do not sum up to one for any finite initial time. Indeed, the weighted sum (7) involves infinitely many "realized" values x_t, even with $t < 0$. The model (1), endowed with adaptive learning, can be reduced to a one-dimensional dynamical system in the space of expected values by inserting (1) inside (5)

$$x_{t+1}^{(e)} = (1-\alpha)\, x_t^{(e)} + \alpha F\left(x_t^{(e)}\right) \tag{8}$$

This means that, given an initial expectation $x_0^{(e)}$, the whole time evolution (or trajectory) of expected values is obtained by the iteration of the one dimensional map

$$g_\alpha(z) = (1-\alpha)\, z + \alpha F(z)\,. \tag{9}$$

Of course, the corresponding time evolution (or trajectory) of realized values $x_t, t \geq 0$, is simply obtained by (1), i.e. by taking the images by F of the expected values, $x_t = F(x_t^{(e)}), t \geq 0$.

The properties of the map (9) are well known (see e.g. Hommes 1994; Chiarella 1988). It is a convex combination of the map F and the identity map, so its graph is included inside the region between the graph of F and the diagonal. This implies that the map g_α and the map F have the same fixed points, i.e. the REEs are fixed points of g_α as well. It is immediate to realize that adaptive expectations are fulfilled for each t, i.e. $x_t^{(e)} = x_t \ \forall t$, if and only if $x_t = F(x_t)$, i.e. at the REE. Instead, the cycles of g_α are in general different from those of F, and the adaptive forecastings are always wrong along invariant sets that are not fixed points of $F(x)$.

Let us assume that the functions are smooth enough on a compact interval of interest, i.e. $F : I \to I$, F of class $C^{(1)}$. It is worth to note that the graph of g_α approaches the graph of F as $\alpha \to 1$, i.e. in the limiting case of naive expectations, whereas the graph of the map g_α approaches the diagonal as $\alpha \to 0$. This implies that for each F a value $\overline{\alpha} \in (0, 1)$ exists such that g_α is an increasing function for any $\alpha \in (0, \overline{\alpha})$ and, as it is well known, an increasing map cannot have cycles of period $k > 1$. In other words, an adaptive learning, with sufficiently low values of α, rules out any dynamic behavior which is more complex than convergence to a REE. However, not all the REEs are stable under adaptive learning. From the properties of the map

g_α, the following well known results follow, which are immediate consequences of the fact that the first derivative of $g_\alpha(z)$ is a convex combination of 1 and $F'(z)$.

(i) If α is sufficiently small $\left(\text{i.e. } \alpha < \min\left\{\frac{1}{1-F'(x)},\ x \in I,\ F'(x) \neq 1\right\}\right)$ then g_α is increasing, so that only REEs exist.
(ii) If a REE x^* is unstable with $F'(x^*) > 1$ then it is also unstable for g_α being $g'_\alpha(x^*) > 1$.
(iii) If a REE x^* is unstable with $F'(x^*) < -1$ then it is stable for g_α for a sufficiently small value of α.

The properties listed above suggest a stabilizing role of adaptive expectations with respect to naive expectations. However, cases in which, for intermediate values of α, dynamic behaviors of the map g_α can be obtained which are more complex than the dynamics of the map F, have been given in the literature (Chiarella 1988; Hommes 1991; Hommes 1994). This happens, for example, with decreasing functions F. In these cases the iteration of F can only exhibit convergence to a fixed point or to cycles of period 2, whereas the corresponding map g_α which governs the time evolution of adaptive expectations, may be noninvertible (a bimodal map) for intermediate values of α, so that cycles of any period and chaotic dynamics can be observed, and even distinct coexisting attractors.

2.2 *Statistical Learning and Reduction to Two Dimensional Dynamics*

Another frequently used learning mechanism is obtained by assuming that, at any time period $t = 0, 1, \ldots$ the agents compute the expected value at the next time period $(t+1)$ as a weighted arithmetic mean of past realized values

$$x_{t+1}^{(e)} = \sum_{k=0}^{t} a_{tk} x_k \tag{10}$$

with weights

$$a_{tk} \geq 0,\ \ k = 0, \ldots, t, \text{ normalized to 1, i.e. } \sum_{k=0}^{t} a_{tk} = 1 \tag{11}$$

Some authors call *statistical learning* this method to obtain expected values (see e.g. Guesnerie and Woodford 1992). The learning mechanism as suggested by Bray (1983) in the form of a simple arithmetic mean

$$x_{t+1}^{(e)} = \frac{1}{t+1} \sum_{k=0}^{t} x_k \tag{12}$$

is a particular case. More general distributions of weights can be proposed, which reflect the different methods that the agents use to exploit information contained in the past observations. These can be obtained by defining, for each time $t \geq 0$, a $(t+1)$-dimensional vector of *relative weights*

$$\omega^{(t)} = \left\{\omega_0^{(t)}, \omega_1^{(t)}, ..., \omega_t^{(t)}\right\} \tag{13}$$

with $\omega_k^{(t)} \geq 0$, from which the weights (11) are computed as

$$a_{tk} = \frac{\omega_k^{(t)}}{W_t}, k = 0, ..., t \qquad \text{with} \qquad W_t = \sum_{k=0}^{t} \omega_k^{(t)} \tag{14}$$

A reasonable assumption for the computation of the relative weights is that old observations are less considered by economic agents, i.e. they use decreasing weights which discount older data (see e.g. Friedman 1979; Radner 1983; Lucas 1986). A simple method to obtain this consist in assigning a fixed value to the weight of the last observed value, say $\omega_t^{(t)} = 1, t \geq 0$, and then the other weights are computed so that the ratio between two successive weights is fixed, that is, $\omega_{k-1}^{(t)}/\omega_k^{(t)} = \rho$, $\rho \in [0, 1]$. With this assumption (13) becomes

$$\omega^{(t)} = \left\{\rho^t, \rho^{t-1}, ..., \rho, 1\right\} \tag{15}$$

i.e. $\omega_k^{(t)} = \rho^{t-k}$, and consequently

$$a_{tk} = \frac{\rho^{t-k}}{W_t} \tag{16}$$

where W_t is the (t-th) partial sum of a geometric series

$$W_t = \sum_{k=0}^{t} \rho^{t-k} = \begin{cases} \frac{1-\rho^{t+1}}{1-\rho} & if \quad 0 \leq \rho < 1 \\ t+1 & if \quad \rho = 1 \end{cases} \tag{17}$$

Statistical learning with "geometrically" distributed weights (16) have been used in Bischi and Naimzada (1997) as a generalization of that proposed by Bray: in fact, for $\rho = 1$ it gives the Bray's average (12). In the other limiting case $\rho = 0$ it reduces to naive expectations $x_{t+1}^{(e)} = x_t$, whereas for intermediate values of the *memory ratio* ρ this learning rule describes agents which, at each time period t, compute their expectations according to a weighted estimation procedure which "exponentially discounts older observations" (see Friedman 1979), that is, an *exponentially fading memory*, see also Foroni et al. (2003), Naimzada and Tramontana (2009), Pecora and Tramontana (2016), Tramontana (2016), and Cavalli and Naimzada (2015), as well as a further generalization with power means in Bischi et al. (2015).

The learning rule (10) with "geometric weights" (16)

$$x_{t+1}^{(e)} = \sum_{k=0}^{t} \frac{\rho^{t-k}}{W_t} x_k \tag{18}$$

can be written as a generalized "adaptive rule" with nonautonomous (i.e. "time-dependent") adjustment speed. In fact,

$$x_{t+1}^{(e)} = \frac{\rho W_{t-1}}{W_t} \sum_{k=0}^{t-1} \frac{\rho^{t-1-k}}{W_{t-1}} x_k + \frac{1}{W_t} x_t = \frac{W_t - 1}{W_t} x_t^{(e)} + \frac{1}{W_t} x_t$$

where the recursive relation

$$W_{t+1} = 1 + \rho W_t, \qquad W_0 = 1. \tag{19}$$

has been used. So, if we define

$$\alpha_t = \frac{1}{W_t}. \tag{20}$$

we get

$$x_{t+1}^{(e)} = (1 - \alpha_t)\, x_t^{(e)} + \alpha_t x_t = (1 - \alpha_t)\, x_t^{(e)} + \alpha_t F\left(x_t^{(e)}\right) \tag{21}$$

which is very similar to an "adaptive rule" (4) except for the fact that the constant speed α is replaced by a time-dependent speed of adjustment given by a decreasing sequence $\{\alpha_t\}$ with $\alpha_t \in (0, 1)$ for each t and $\alpha_t \to (1 - \rho)$ as $t \to +\infty$. Hence, for $t \to +\infty$, the nonautonomous recurrence (21) tends to the *limiting form*

$$x_{t+1}^{(e)} = g_{1-\rho}(x_t^{(e)}) = \rho x_t^{(e)} + (1 - \rho) F(x_t^{(e)}). \tag{22}$$

i.e., in the long run it behaves like a model with a standard adaptive rule, with speed of adjustment $\alpha = 1 - \rho$. This fact led many authors to consider the two learning rules, the adaptive rule (4) and the statistical rule (18), as practically equivalent, and justify this equivalence statement by the property that the dynamics of the expected values under both the learning rules are governed, in the long run, by the a one-dimensional map which has the same form $g_{1-\rho}(z)$, given by (9). However, even if the limiting sets are the same, their time evolutions are different, because starting from the same initial conditions, the two learning mechanisms exhibit different transient dynamics due to the fact that during the early iterates the dynamics with statistical learning (18), governed by (21), is different from the one governed by (22), and in the presence of several attractors this may be crucial to decide which one will be reached in the long run. In particular, if several attractors are present, convergence to different attractors under the two learning mechanisms may be observed even starting from the same initial condition, thus giving different equilibrium selection results.

This can be equivalently stated by saying that *even if the two different kinds of learning dynamics have the same attracting sets, their basins of attraction may be different.* This is also true in the particular case of Bray learning (12), corresponding to the limiting case $\rho = 1$, as we shall see in the following through some examples.

Some remarks on the crucial role of initial conditions in models with Bray's learning can be found in the literature. For example, in Holmes and Manning (1988) the learning rule (12) is used in a nonlinear cobweb model with decreasing F, and it is stressed that such type of learning has a stabilizing effect on the long run dynamics. However, the authors remark that the short and intermediate run dynamics can be rather complex and of considerable interest. A similar argument is given in Dimitri (1988) where a quadratic map F is proposed as a modification of a linear model of price dynamics with $p_t^{(e)}$ computed according to (12) as proposed in Bray (1983). On the basis of numerical results Dimitri writes "...the evolution of the model is indeed very much dependent upon the starting position..." as a comment to the fact that even if a REE is locally stable, divergent price sequences are obtained even if initial conditions are taken rather close to the REE. These considerations lead us to face the problem of the basins of attraction. This is not, in general, an easy task for nonautonomous recurrences like (21), because for nonautonomous recurrences the ω-limit sets are not invariant sets, due to the fact that the iterated map changes as t varies. However, a global characterization of the basins is possible for (21) since it can be reduced to an autonomous two-dimensional map. This is easily obtained by noticing that, from (19), the sequence $\{\alpha_t\}$ defined in (20) can be defined recursively as

$$\alpha_{t+1} = \frac{\alpha_t}{\alpha_t + \rho}, \quad \alpha_0 = 1$$

So, the model (1) endowed with learning (18) can be written as

$$\begin{cases} x_t = F\left(x_t^{(e)}\right) \\ x_{t+1}^{(e)} = (1 - \alpha_t)\, x_t^{(e)} + \alpha_t x_t \\ \alpha_{t+1} = \frac{\alpha_t}{\alpha_t + \rho} \end{cases}$$

with initial conditions x_0 (the initial realized value) and $\alpha_0 = 1$. This recursive relation is already known in the literature, at least for the limiting case of Bray learning, i.e. for $\rho = 1$ (see e.g. Marimon 1997). Following the same procedure as in the case of adaptive expectations, we can use (1) to obtain a mapping (which is two-dimensional in this case) which defines the time evolution of the expected values. However, it is important to remark that in this case the iteration procedure starts with the value observed in $t = 0$, given by x_0, and this implies that the first expected value used to start (23) is given, according to (10), by $x_1^{(e)} = x_0$. So, the sequences of expected values generated by (21) can be obtained from the iteration of (23) starting from $\alpha_0 = 1$ and $x_1^{(e)} = x_0$. This means that the difference equation by

which $x_t^{(e)}$ is recursively computed is shifted of one period with respect to the other one:

$$\begin{cases} x_{t+2}^{(e)} = (1-\alpha_{t+1})\, x_{t+1}^{(e)} + \alpha_{t+1} F\left(x_{t+1}^{(e)}\right) \\ \alpha_{t+1} = \frac{\alpha_t}{\alpha_t + \rho} \end{cases} \tag{23}$$

Following Bischi and Naimzada (1997) and Bischi and Gardini (1997), in order to study the general properties of the two-dimensional map (23) we rewrite it in the equivalent form

$$T : \begin{cases} z_{t+1} = \frac{\rho W_t}{1+\rho W_t} z_t + \frac{1}{1+\rho W_t} F(z_t) \\ W_{t+1} = 1 + \rho W_t \end{cases} \tag{24}$$

where $z_t = x_{t+1}^{(e)}$ and W_t is defined in (17). The sequence of expected values of the model (1) with learning (18) are obtained from the trajectories of the two-dimensional recurrence (24) provided that the conditions are chosen as:

$$z_0 = x_1^{(e)} = x_0 \quad \text{and} \quad W_0 = 1 \tag{25}$$

Starting from a given (z_0, W_0) the iterations of the map T uniquely defines the trajectory $\tau = \left\{(z_t, W_t) = T^t(z_0, W_0), t \geq 0\right\}$ and if $(z_0, W_0) = (x_0, 1)$ then the sequence $\{z_t, t \geq 0\}$ represents the time evolution of the expected variables $\left\{x_t^{(e)}, t \geq 1\right\}$ from which the sequence of realized values $\{x_t\}$ starting with the given x_0 is simply obtained as the images under the function F:

$$x_t = F\left(x_t^{(e)}\right) \qquad t \geq 1 \tag{26}$$

In other words, if $\{(z_0, W_0), (z_1, W_1), \ldots, (z_t, W_t), \ldots\}$ is the sequence generated by the map T starting from the initial condition $(z_0, W_0) = (x_0, 1)$, then $\left\{x_1^{(e)} = z_0, x_2^{(e)} = z_1, \ldots, x_{t+1}^{(e)} = z_t, \ldots\right\}$ is the sequence of expected values, and $\{x_0, x_1 = F(z_0), \ldots, x_t = F(z_{t-1}), \ldots\}$ is the corresponding sequence of realized values. Thus the study of the general model (1) with learning rule (18) is reduced to that of a two-dimensional map with initial conditions constrained on the line $W = 1$ (*line of initial conditions*). The class of maps (24) has been initially studied in Bischi and Gardini (1997), which inspired a stream of literature on maps with vanishing denominator, see Bischi et al. (1999), Bischi et al. (2003), and Bischi et al. (2005), from which several applications followed, e.g. Tramontana (2016) and Gu and Hao (2007).

3 Limit Sets and Basins of Attraction for Statistical Learning With fading Memory

Any trajectory of (24) starting from initial conditions on the line $W = 1$ is confined in the strip $0 < W < \frac{1}{1-\rho}$. In fact, this strip is mapped into itself by T because the second difference equation in (24), which gives the dynamics of the variable W, is independent of z and gives a monotonically increasing sequence (the partial sums of a geometric series of ratio ρ) and if $0 \le \rho < 1$ such sequence $\{W_t\}$ converges to the sum of the geometric series

$$W^* = \frac{1}{1-\rho} \,. \tag{27}$$

For $0 \le \rho < 1$ the line $W = W^*$ is an invariant and globally attracting line for the map T, on which the ω-limit sets of all its trajectories are located. For this reason we shall call it *line of ω-limit sets* . The restriction of T to this line is given by the one-dimensional map

$$g(z) = \rho z + (1-\rho)F(z) \,, \tag{28}$$

already obtained in (22) as the limiting form of the nonautonomous recurrence (21). The map (28) will be called *limiting map*, since it governs the asymptotic behavior of the map T. This implies, as proved in Bischi and Gardini (1996), that any k-cycle $A = \{z_1^*, \ldots, z_k^*\}$ of the map $g_\rho(z)$ is in one-to-one correspondence with a k-cycle $\mathcal{A} = A \times \{W^*\} = \{(z_1^*, W^*), \ldots, (z_k^*, W^*)\}$ of the map T, located on the line of ω-limit sets. Moreover, the attractors of the model (1) with learning scheme (18), as well as their basins of attraction, can be studied on the basis of the following proposition, given in Bischi and Gardini (1996) (see also Bischi and Naimzada 1997):

Proposition 1 *Let A be a k-cycle, $k \ge 1$, of the map $g_{1-\rho}(z)$, $0 \le \rho < 1$. Then*

(i) if A is attracting for the limiting map $g_\rho(z)$, then the set $\mathcal{A}=A \times \{W^*\}$ is an attracting cycle of the map T, and $F(A)$ is an attracting cycle of the model (1) with learning scheme (18);

(ii) the basin of attraction D_1 of the attractor $F(A)$ of the model (1) with learning scheme (18) is given by the intersection of the two-dimensional basin $\mathcal{B}$ of the cycle $\mathcal{A}$ of the map T with the line of initial conditions $W = 1$.

We recall that the case $k = 1$ corresponds to a fixed point z^* of $g(z)$, and $F(A) = F(z^*) = z^*$ is a REE, since the fixed points of $g(z)$ are also fixed points of $F(z)$.

The part (i) of the Proposition 1 confirms that the asymptotic behavior, i.e. the kinds of attractors and their stability properties, are the same as those of a standard adaptive learning rule with adaptive coefficient $\alpha = 1 - \rho$. For example, a sufficient condition for the attractivity of a REE z^*, under learning (18) with $\rho < 1$, is given by $\left|g'(z^*)\right| < 1$, that is,

$$-\frac{\rho+1}{1-\rho} < F'(z^*) < 1 \,. \tag{29}$$

However, the most important implications of Proposition 1 are due to part (ii), since it suggests a general procedure to obtain the boundaries of the basins of attraction when two or more coexisting attractors are present, as often occurs in the case of nonlinear models. This is an important issue that cannot be studied on the basis of the limiting map g_ρ, because the initial conditions are to be taken on the line $W = 1$, whereas g_ρ only governs the dynamics near the line of ω-limit sets $W = W^*$. This means that only a global knowledge of the two-dimensional map T allows one to follow the short-run behavior, during which the dynamics is not governed by the limiting map g.

Moreover, as outlined in the Proposition 1, the basins of attraction of the two-dimensional map T, whose intersection with the line of initial conditions $W = 1$ gives the basins of the model with statistical learning, are obtained by considering the preimages of proper neighborhoods of the attracting sets located along the line of ω-limit sets. We recall that the two dimensional *basin of attraction* of an attractor A of the map T is the open set of points which generate trajectories converging to A:

$$\mathcal{B}(A) = \left\{(z, W) \,|\, T^t(z, W) \to A \text{ as } t \to +\infty\right\} \quad . \tag{30}$$

A closed invariant set $A \subset \{W = W^*\}$ is called *asymptotically stable* (or *attracting*) if a neighborhood U of A exists such that $T(U) \subseteq U$ and $T^n(x) \to A$ as $n \to +\infty$ for each $x \in U$. Then, the *basin* of A is obtained by taking all the preimages of the points of U

$$\mathcal{B}(A) = \bigcup_{n=0}^{\infty} T^{-n}(U)$$

where $T^{-n}(x)$ denotes the set of all the preimages of x of rank n, i.e. the set of all the points which are mapped into x after n iterations of T. So, the study of the two-dimensional basin is based on the study of the inverses of T. In the case of the map (24) we have that the properties and the qualitative changes of its basins are strongly influenced by the presence of the denominator which can vanish along the line $W = -\frac{1}{\rho}$ and, in particular, by the points in which the first component of T assumes the form 0/0, see Bischi et al. (1999, 2003, 2005), Gardini et al. (2007), and Bischi and Gardini (1997) for the particular class of triangular maps (24). In these papers it is proved that the presence of points where a component of the map assumes the form 0/0, called *focal points*, may have important consequences on the structure of the basins and their global bifurcations, because fans of basins boundaries arise from them giving peculiar finger-shaped structures called *lobes*. The existence of lobes, originating from the focal points, may have important consequences on the structure of the basins of attraction of the model with learning (18) whenever they intersect the line of initial conditions $W = 1$. This occurrence causes the creation of basins with a complicated topological structure, such as basins formed by many disjoint intervals, as we shall see in the next section.

3.1 Global Properties and Structure of the Basins of The two-Dimensional Triangular Map

In the following we briefly recall some definitions and properties specific to maps with vanishing denominator (see Bischi et al. 1999, 2003, 2005 for a more complete treatment). Let us consider a map $(x, y) \to (x', y') = T(x, y)$ of the form

$$T : \begin{cases} x' = F(x, y) \\ y' = G(x, y) \end{cases} \tag{31}$$

where x and y are real variables and at least one of the components has the form of a fractional rational function, i.e.

$$F(x, y) = \frac{N_1(x, y)}{D_1(x, y)} \quad \text{and/or} \quad G(x, y) = \frac{N_2(x, y)}{D_2(x, y)} \tag{32}$$

The functions $N_i(x, y)$ and $D_i(x, y)$, $i = 1, 2$, are defined in the whole plane $\mathbb{R}^2$, so that the set with no definition δ_s of the map T coincides with the locus of points in which at least one denominator vanishes:

$$\delta_s = \left\{(x, y) \in \mathbb{R}^2 | D_1(x, y) = 0 \text{ or } D_2(x, y) = 0\right\} \tag{33}$$

The two dimensional recurrence obtained by the iteration of T is well defined, i.e. it generates not terminating trajectories, provided that the initial condition belongs to the set E given by

$$E = \mathbb{R}^2 \setminus \bigcup_{k=0}^{\infty} T^{-k}(\delta_s) \tag{34}$$

so that $T : E \to E$. We recall here the following definition

Definition *A point $Q = (x_Q, y_Q)$ is a focal point if at least one component of the map T takes the form 0/0 in Q and there exist smooth simple arcs $\gamma(\tau)$, with $\gamma(0) = Q$, such that $\lim_{\tau \to 0} T(\gamma(\tau))$ is finite. The set of all such finite values, obtained by taking all the arcs $\gamma(\tau)$ through Q, is the prefocal set δ_Q.*

Roughly speaking, a *prefocal curve* is a set of points for which at least one inverse exists which maps (or "focalizes") the whole set into a single point, called *focal point*. For maps with a vanishing denominator, new kinds of contact bifurcations have been evidenced which involve the singularities defined above. In particular, contacts between basin boundaries and prefocal curves may cause the creation of particular structures of the basin boundaries, denoted as *lobes* and *crescents*. These particular structures have been observed in the study of discrete dynamical systems of the plane which arise in some different contexts, see e.g. Billings and Curry (1996), Billings et al. (1997), Foroni et al. (2003), Gardini et al. (1999), Tramontana (2016),

Pecora and Tramontana (2016), Cavalli and Naimzada (2015), and Yee and Sweby (1994). As already mentioned, the existence of lobes, originating from the focal points, has important consequences on the structure of the basins of attraction of the model with statistical learning whenever they intersect the line of initial conditions $W = 1$, causing the creation of one dimensional basins, of the model (1) with learning (18) with a complicated topological structure.

We now briefly describe the basic mechanism leading to the formation of lobes and crescents. In order to do this, let us consider the map T given in (24), which we rewrite as $T : (z, W) \to (z', W')$, i.e.

$$T : \begin{cases} z' = \frac{\rho W}{1+\rho W} z + \frac{1}{1+\rho W} F(z) \\ W' = 1 + \rho W \end{cases} \tag{35}$$

and we consider the image of an arc crossing through a *focal point*. We shall see that, according to the general results given in Bischi et al. (1999), a one-to-one correspondence is obtained between the slopes of the arcs through a focal point and the points in which their images cross the corresponding prefocal curve.

We first notice that the map (35) is not defined in the whole plane, because the denominator of the first component vanishes on the points of the line δ_s of equation $W = -\frac{1}{\rho}$. So, in order to have a well defined recurrence we must exclude from the phase plane of T the singular line as well as all its preimages of any rank δ_s^{-n} for each $n \geq 1$, belonging to a sequence of lines located below the singular line obtained by backward iteration of the second component of T, i.e.

$$W = \frac{W' - 1}{\rho}. \tag{36}$$

So, δ_s^{-1} has equation $W = -\frac{1+\rho}{\rho^2}$ and is located below δ_s, and analogously δ_s^{-n}, the set of points which are mapped in the singular line after n iterations of T, are located on the line of equation

$$W = -\frac{\sum_{k=0}^{n} \rho^k}{\rho^{n+1}} = -\frac{1 - \rho^{n+1}}{\rho^{n+1} - \rho^{n+2}}. \tag{37}$$

and the phase space of the recurrence defined by the map T is given by

$$E = \mathbb{R}^2 \setminus \bigcup_{n=0}^{\infty} \delta_s^{-n}. \tag{38}$$

where $\delta_s^{-(n+1)}$ is below δ_s^{-n} for each $n \geq 1$. In the map (35) only the first component has denominator, which vanishes in the points of the singular line $y = -\frac{1}{\rho}$, where the numerator becomes $F(x) - x$, hence it vanishes at every fixed point of the function F

(and thus also of the limiting map (22)). It follows that the a focal point is necessarily of type $\left(x^*, -\frac{1}{\rho}\right)$, where x^* is a fixed point of $F(x)$.

In order to explain the role of a *focal point* and the related *prefocal set* in the geometric and dynamic properties of the map T, following the arguments given in Bischi et al. (1999) we consider a smooth simple arc γ transverse to δ_s and how it is transformed by T. Let $(z_0,-1/\rho)$ be the point where γ intersects δ_s and assume that the arc γ is deprived of $(z_0,-1/\rho)$. If $z_0 \neq x^*$, i.e. $F(z_0) \neq z_0$, then the image $T(\gamma)$ is made up of two disjoint unbounded arcs asymptotic to the line of equation $y = 0$, as qualitatively shown in Fig. 1. A different situation may occur if $z_0 = x^*$, i.e. $F(z_0) = z_0$, because in this case the numerator of the first component also vanishes, and the limit of $T(\gamma)$ may take finite values as $(z, W) \to \left(x^*, -\frac{1}{\rho}\right)$, so that $T(\gamma)$ is a bounded arc (as qualitatively sketched in Fig. 1 for the arc γ_2). If m is the slope of the tangent to the smooth arc γ in the focal point $Q = \left(x^*, -\frac{1}{\rho}\right)$ then in Bischi et al. (1999) it is proved that the image $T(\gamma)$ crosses the line $W = 0$ in the point $(u_m, 0)$ with

$$u_m(x^*) = x^* + \frac{F'(x^*) - 1}{\rho m}. \tag{39}$$

This means that the images of the arcs crossing through $\left(x^*, -\frac{1}{\rho}\right)$ with slope $m \neq 0$ are bounded arcs (as qualitatively shown in the right panel of Fig. 1), and as m varies in $\mathbb{R}$ all the points of the line $W = 0$ are obtained, provided that $F'(x^*) \neq 1$. Thus the line of equation $W = 0$ represents the *prefocal set* δ_Q for the map (35). The situation in which $F'(x^*) = 1$ can be considered as a bifurcation case (see Bischi et al. 2005).

This suggests some consequences when we consider the preimages. The map (35) may be a noninvertible map, because the number of distinct inverses of T depends on the function $F(x)$. In fact, even if a point (z, W) has a unique image under the

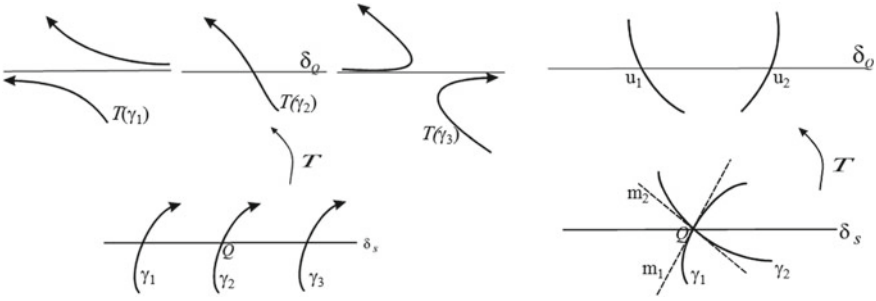

Fig. 1 Schematic picture of the action of a two-dimensional map on an arc crossing a singular curve δ_S along wich a denominator vanishes. Left: The arcs γ_1 and γ_3 cross the singular curve in a generic point of δ_S whereas γ_2 crosses it through a focal point. Right: Two arcs crossing δ_S through a focal point with different slopes are mapped into finite arcs crossing the prefocal curve δ_Q in different points

application of T, given by $\left(z', W'\right) = T\,(z, W)$, the backward iteration of T may not be uniquely defined, since given a point $\left(z', W'\right)$ its preimages (z, W) are obtained by solving, with respect to the unknowns z and W, a system which may have several real solutions, i.e. several inverses. If n is the number of distinct inverses we denote them by $T_i^{-1}\left(z', W'\right)$ for $i = 1, ..., n$ and $T^{-1}\left(z', W'\right) = \bigcup_{i=1}^{n} T_i^{-1}\left(z', W'\right)$. Moreover, if $F(x)$ has N fixed points (hence also T has N fixed points) then the prefocal line must belong to a region, say Z_N, whose points have N distinct rank-1 preimages.

To sum up, *for each focal point* $Q_i = \left(x_i^*, -\frac{1}{\rho}\right)$ *the map T in* (35) *defines a one-to-one correspondence between the slope m of an arc* γ *through* Q_i *and the point* $(u, 0)$ *in which the image* $T(\gamma)$ *crosses the prefocal curve* δ_Q, *given by*

$$\begin{array}{lll} m \to (u,0) & : & u = x_i^* + \frac{F'(x_i^*)-1}{\rho m} \\ (u,0) \to m & : & m = \frac{F'(x_i^*)-1}{\rho\left(u-x_i^*\right)} \end{array} \tag{40}$$

Some consequences of this correspondence, important for the characterization of the basins' boundaries and their bifurcations, are deduced by considering a smooth arc ω that intersects the prefocal line in two points. In this case, the N rank-1 preimages of ω, say $T_i^{-1}(\omega)$, $i = 1, ..., N$, are arcs such that each $T_i^{-1}(\omega)$ has a loop with knot in the focal point $Q_i = (x_i^*, -\frac{1}{\rho})$. This implies that a remarkable contact bifurcation occurs when a smooth curve segment ω moves towards the prefocal curve δ_Q until it has a contact and then crosses δ_Q (as qualitatively shown in Fig. 2). As ω moves toward δ_Q, its preimages move towards Q_i, and when ω becomes tangent to δ_Q then each preimage $\omega_{-1}^i = T_i^{-1}(\omega)$ has a cusp point at Q_i. The slope of the common tangent line to the two arcs that join in Q_i is given by $m_i(u_c)$, according to (40). If the curve segment ω moves further, so that it crosses δ_Q in two points, say $(u_1, 0)$ and $(u_2, 0)$, then its preimages form loops with double points at the focal points Q_i.

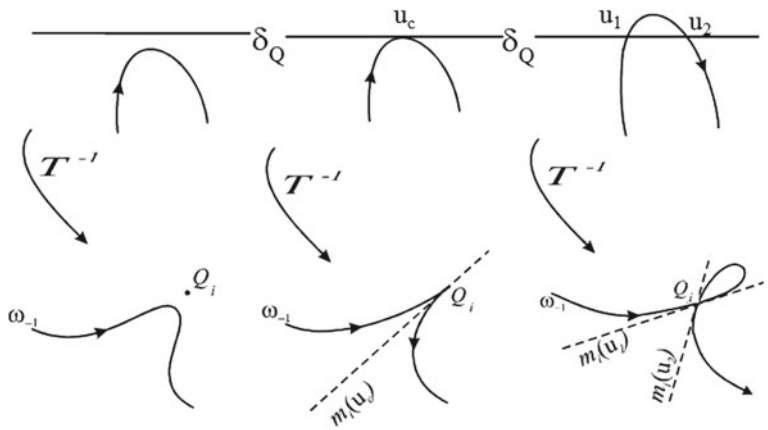

Fig. 2 Qualitative picture of a preimage of an arc moving towards a prefocal line until crossing it

This kind of contact bifurcation is important in the study of the boundaries of the basins of attraction, because if ω is a portion of a basin boundary, a contact between ω and δ_Q implies that a loop is created along the basin boundary, because a basin boundary is backward invariant, i.e. it includes all the preimages of any portion of it, and the portion of the basin inside the loop is a lobe, as we shall discuss in the next sections. Moreover, in the case of noninvertible maps the creation of *crescents* can be obtained as well, obtained from the merging of lobes as qualitatively shown in Fig. 3. It is caused by contacts of a critical curve LC with a basin boundary which already includes lobes which merge along LC_{-1} after the contact (see e.g. Mira et al. 1996 for a definition of critical curves).

Now, let us consider the forward iteration of T. It is easy to see that the image of rank-n of the prefocal line $W = 0$ belongs to the line of equation $W = W_n$ where

$$W_n = \frac{1 - \rho^{n+1}}{1 - \rho} \tag{41}$$

i.e. a sequence of lines parallel to the prefocal line δ_Q and convergent to the line of the ω-limit sets $W = W^*$. This implies that any cycle belonging to the ω-limit set $W = W^*$ is transversely attracting. This property is important in order to study the boundaries of the basins. In fact, recall that, in general, the boundaries of a basin are obtained by taking the stable sets of some cycles on it. In the case of maps (35) such cycles can only be located on the line of ω-limit sets. To get the stable set W^s of a

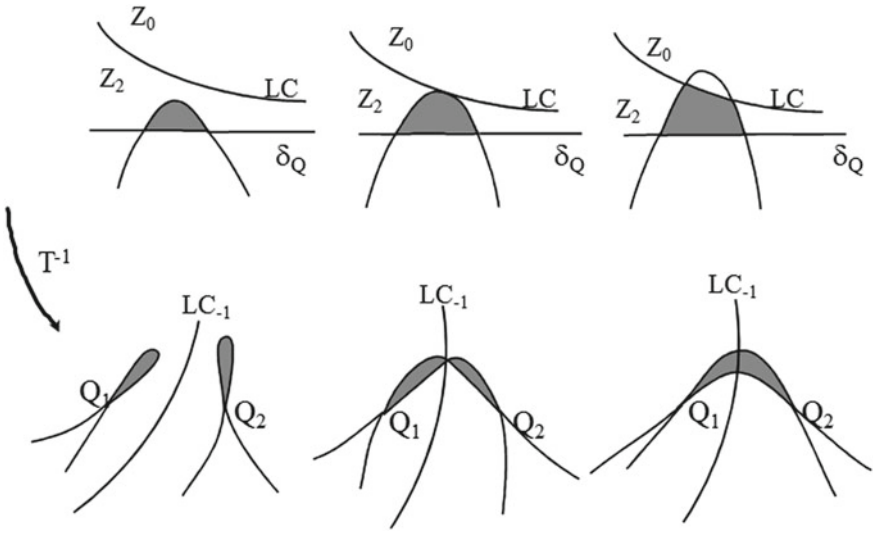

Fig. 3 Qualitative sketch to describe the formation of a *crescent* obtained by the merging of two *lobes* when a portion of a basin of attraction crossing the prefocal line δ_Q has a contact with a critical curve LC

saddle it is enough to take the preimages of any rank of a local stable set W^s_{loc}, that is $W^s = \bigcup_{n=0}^{\infty} T^{-n}(W^s_{loc})$, where W^s_{loc} is transverse to the line $W = W^*$. Due to the expansive character of T^{-1} along the W direction, as defined in (36), such preimages must necessarily reach, in a finite number of steps, the prefocal line $W = 0$. So, all these preimages must necessarily cross the singular line $W = -\frac{1}{\rho}$ through focal points Q_i.

From this observation it follows that the stable set of any saddle cycle of T, obtained by taking the preimages of a local stable set, is made up of branches issuing from the focal points. In fact, the preimages of any local stable set, transverse to the line of ω-limit sets $W = W^*$, necessarily go back to the prefocal line $W = 0$ in a finite number of steps. Thus any stable set must be made up of branches which "cross" the singular line in the focal points, i.e. are "focalized" through the focal points. This argument, applied to the stable set of some saddle cycle belonging to the line of ω-limit sets, constitutes the global mechanism which causes the formation of the particular structures of the basins which will be shown in the examples.

3.2 Increasing Maps

We have seen that even if the law of motion (1) with the two different kinds of learning (4) and (18) has the same attracting sets, the corresponding basins of attraction are generally different. This is related to the fact that the asymptotic dynamics obtained with both the learning mechanisms are governed by the map $g_\rho = g_{1-\alpha}$, whereas the basins are obtained by different procedures.

However, things are simpler when F is an increasing map. In this case the limiting map g is also increasing. As it is well known, for a continuous and increasing map the only invariant sets are the fixed points, and when several fixed points exist, say $x_1^* < x_2^* < \ldots < x_k^*$ they are alternatingly stable and unstable: the unstable fixed points are the boundaries that separate the basins of the stable ones. Indeed, in this case the same situation also holds for the basins of the stable fixed points under the statistical learning.

Proposition *If F is increasing then the basins with adaptive learning and speed α are the same as those for the statistical learning with fading memory and ratio $\rho = 1 - \alpha$.*

Proof Let F be a continuous and increasing function and let x^* be an unstable fixed point of F. The stable set of the saddle $S = (x^*, 1/(1-\rho))$ is obtained by taking the preimages of any rank of the local stable set of S, which is included in the line $z = x^*$. The rank-1 preimages of a point $\left(x', W'\right)$ are obtained by

$$\begin{cases} F(z) + z\left(W' - 1\right) = x'W' \\ W = \frac{W'-1}{\rho} \end{cases}$$

and for any $W' \geq 1$ this has only one solution, because the left hand side of the first equation is increasing if F is increasing and $W' \geq 1$. In particular, the only rank-1 preimage of a point (x^*, W') belongs to the line $z = x^*$, so that the projections of the unstable fixed points on the line of initial conditions $W = 1$ are the only boundaries which separate the basins, like in the one-dimensional map g □.

We remark that two dimensional basins of the attracting fixed points of the map T, located on the line $W = W^*$, may include portions which do no belong to the vertical lines through the saddle points, but these are necessarily confined in the region $W < 1$, so that they have no influence on the basins of the model (1) with learning (18). An example will be shown below.

3.3 The Particular Case of Bray Learning

We have seen that the case of the statistical learning (12) proposed by Bray (1983) can be obtained from the statistical learning with fading memory (18) in the limiting case $\rho \to 1^-$. It can be noticed that in this case any REE x^* with $F'(x^*) < 1$ is stable, i.e. it can be "learned" by the agents, because the stability condition (29) is always satisfied as $\rho \to 1^-$. In other words, for the general model (1) with Bray's learning (12), the steady states x^* characterized by $F'(x^*) < 1$ are locally attracting equilibria, whereas those with $F'(x^*) > 1$ are unstable saddles. This confirms, and extends, the stability results obtained, for particular models, by Bray (1983), Dimitri (1988), and Holmes and Manning (1988). That is: *in the case of Bray's learning* (12) *any complexity is lost, and any trajectory is either divergent or convergent to a stable REE* x^*.

However, besides divergent trajectories there may be two or more coexisting stable REEs, and the arguments given above about the complexity in the basins also hold in this case. In fact, even if $W^* = 1/(1-\rho) \to +\infty$, for each REE x^* with $F'(z^*) < 1$ the invariant lines $z = x^*$ are attracting sets whose basins can be obtained following the same procedure outlined in the previous sections. The triangular map (35) becomes

$$T : \begin{cases} z' = z + \frac{F(z)-z}{1+W} \\ W' = 1 + W \end{cases}$$

with initial condition taken on the line $W = 1$. The line with no definition is $W = -1$, on it a focal point is associated with each REE and the prefocal line is $W = 0$.

Except for the uninteresting case in which a unique unstable REE exists (and all the trajectories are divergent), or the simple case in which only one globally attracting REE exists, there are both attracting lines $z = x_i^*$ and repelling lines $z = z_j^*$, associated with stable REE x_i^* and unstable REE z_j^* respectively. Then, any repelling invariant line $z = z_j^*$ has a stable set (made up of all the preimages of any rank of the line $z = z_j^*$) which separates different basins. Such preimages may intersect the

prefocal line of T. Thus, depending on the topological structure of these preimages, associated with the inverses of T, the basins on the line $W = 1$ may have a simple or complex geometrical structure.

4 Examples

4.1 Unimodal Maps: A Cobweb Model

One of the simplest models expressed by a law of motion of the form $x_t = F\left(x_t^{(e)}\right)$ is the *cobweb model* (see e.g. Nerlove 1958, Hommes 1991, Chiarella 1988, Jensen and Urban 1984). In the market of a given good, let $q_D = D(p)$ and $q_S = S(p)$ be the demand and supply functions respectively. At the time t, q_D depends on the current price p_t, whereas q_S depends of the price $p_t^{(e)}$ expected by producers at the previous time in which they decided their production. If the production delay is taken as the time unit, the market clearing condition becomes $D\left(p_t\right) = S\left(p_t^{(e)}\right)$, and assuming that $D(p)$ is a continuous and decreasing function (hence invertible) the law of motion of the market clearing price is $p_t = D^{-1}S(p_t^{(e)})$ at which the adaptive learning (4) or the statistical learning (18) can be applied. As an exercise to illustrate the results of the previous sections, and to compare the two kinds of expectations, we consider a cobweb model where $F(x) = D^{-1}S(x)$ is a quadratic map, like in Jensen and Urban (1984) or Dimitri (1988), where a linear demand function is considered together with a backward bending supply curve, expressed by a quadratic function, so that $F(x)$ is conjugate to the standard logistic map

$$f(x) = \mu x\,(1 - x)\,,\ \mu > 1 \tag{42}$$

So, in the following we consider a model $x_t = f\left(x_t^{(e)}\right)$ with f given by (42). In this case, the dynamics of the expected prices under the assumption of adaptive expectations (4) is governed by a quadratic map as well, given by

$$z' = g_\alpha(z) = (1 - \alpha)\,z + \alpha\mu z\,(1 - z) \tag{43}$$

whereas under the assumption of statistical learning (18) the dynamics of the expected prices is obtained by the two-dimensional map (24)

$$T : \begin{cases} z_{t+1} = \frac{\rho W_t z_t + \mu z_t(1-z_t)}{1+\rho W_t} \\ W_{t+1} = 1 + \rho W_t \end{cases} \tag{44}$$

and the limiting map which governs the asymptotic behavior is:

$$z' = g_{1-\rho}(z) = \rho z + (1-\rho)\,\mu z\,(1-z) \tag{45}$$

For each $\mu > 1$ there are two non negative REEs, given by

$$s^* = 0 \qquad \text{and} \qquad p^* = \frac{\mu - 1}{\mu}$$

where s^* is an unstable fixed point of f, with $f'(s^*) > 1$, hence it is also an unstable fixed point of g for each $\rho \in [0, 1)$, whereas p^* is stable for $1 < \mu \leq 3$, and unstable for $\mu > 3$ with $f'(p^*) < -1$. This means that the REE p^* is stable under the assumption of adaptive expectations provided that α is sufficiently small (and the same is true for the statistical learning with ρ sufficiently close to 1). In the following we shall consider values of α or ρ such that the REE p^* is stable, however even in this case, p^* is not globally stable since for each $\alpha \geq 0$ ($\rho \leq 1$) diverging sequences of expected values can be obtained as well. This raises the question of the study of the basins of attraction, i.e. the delimitation of the boundary which separates the set of initial conditions that generate trajectories converging to the REE (i.e. the basin of attraction of p^*) from the set of initial conditions that generate unbounded trajectories (i.e. the basin of attraction of infinity). This question is very easily solved for the case of adaptive learning, whose study requires a simple analysis of the one-dimensional quadratic map (43), for which the basin of p^* is given by the interval $]0, g_\alpha^{-1}(0)[=]0, 1[$. Instead, for the statistical learning a global analysis of the two dimensional map (44) requires more advanced methods. In fact, (44) is a noninvertible map with two focal points,

$$Q_1 = \left(0, -\frac{1}{\rho}\right) \quad \text{and} \quad Q_2 = \left(\frac{\mu - 1}{\mu}, -\frac{1}{\rho}\right). \tag{46}$$

In Fig. 4a, obtained with $\rho = 0.75$ and $\mu = 6$, the basins of the two-dimensional map (44) are shown: The white region represents the set of points converging to the fixed point $P = (p^*, W^*) = \left(\frac{\mu-1}{\mu}, \frac{1}{1-\rho}\right)$, and the grey region represents the set of points which generate diverging trajectories. The intersections with the line of initial conditions $W = 1$ represent the respective one-dimensional basins of the cobweb model with statistical learning (18) given by the interval $(0, \overline{z})$ with $\overline{z} = 1.125$. Instead, if we consider the adaptive learning (4) with $\alpha = 0.25$, so that the dynamics of expected prices are governed by the same one-dimensional map, the basin of the REE p^*, is $\left(0, g_{0.25}^{-1}(0)\right) = (0, 1.5)$, where $g_{0.25}^{-1}(0)$ is the preimage of the unstable fixed point s^* different from itself. This basin coincides with the portion of the line of ω-limit sets $W = W^*$ included in the white region of Fig. 4. A trajectory starting from $x_0^{(e)} = 1.2$ converges to p^* under adaptive expectations, whereas for the model with statistical learning (18) with $\rho = 0.75$, for which the asymptotic dynamics of the expected prices are governed by he same limiting map, the trajectory starting

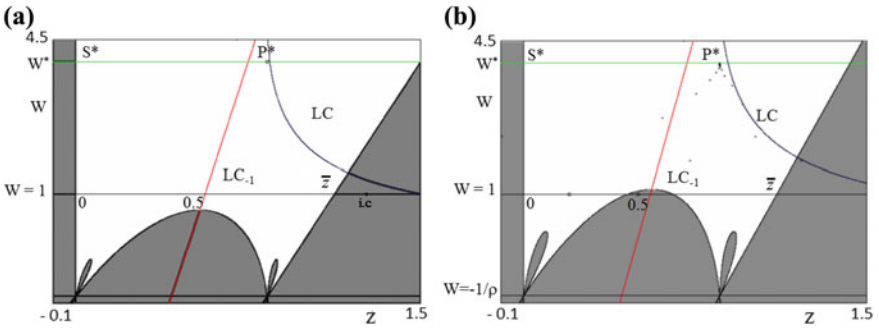

Fig. 4 Numerically generated basins of attraction of the two-dimensional map (44): The white region represents the set of points converging to the fixed point P^* and the grey region represents the set of points which generate diverging trajectories. **a** $\rho = 0.75$ and $\mu = 6$; **b** $\rho = 0.75$ and $\mu = 7$

from $z_0 = p_1^{(e)} = 1.2$ diverges. So, with the same starting condition, the trajectory obtained with the standard adaptive learning converges to the REE p^*, whereas the model with statistical learning (18) with $\rho = 1 - \alpha$ does not converge.

An even more evident difference is obtained in the situation shown in Fig. 4b, obtained with $\rho = 0.75$ and $\mu = 7$. Now the basin $B(p^*)$ is formed by two disjoint intervals, because a "hole" formed by points which generate diverging trajectories is nested inside $B(p^*)$.

It can be noticed that the size, in the z direction, of $\mathcal{B}(\mathcal{P})$, increases for higher values of W, so that stronger shocks are necessary to bring the phase point inside $\mathcal{B}(\infty)$, i.e. the system is less vulnerable with respect to exogenous perturbations as time goes on. Loosely speaking we may say that as the amount of information (i.e. the number of observed realized values) increases the system has a greater probability to converge, because agents learn to behave more and more rationally as time goes on.

The above considerations are even more evident when applied to situations like the one shown in Fig. 5, obtained for $\rho = 0.75$ and $\mu = 9$. In this case the basin of p^* is formed by 4 disjoint intervals, due to the presence of lobes of $\mathcal{B}(\infty)$ intersecting the line of initial conditions $W = 1$.

We now describe a procedure to obtain the exact delimitation of the boundary $\mathcal{F}$ that separates the two basins. In fact, the complementary set of $\mathcal{B}(\infty)$ is the set of bounded trajectories which converge to invariant sets of the limiting map g, on the line $W = W^*$. As remarked above, the attractor always coincides with the REE if the memory ratio ρ is sufficiently close to 1, whereas for lower values of ρ the bounded attractor of the map $g_\rho(z)$ may be a cycle or even a chaotic set. Here we are only interested in parameters' values for which the REE p^* is locally stable, but it is clear that arguments similar to those used below hold independently of the topological structure of the attracting set of $g_\rho(z)$. Thus $P = (p^*, W^*)$ denotes the attractor of T located on the line $W = W^*$, whose basin will be denoted by $\mathcal{B}(\mathcal{P})$. The frontier $\mathcal{F}$ behaves as a repelling set for the points near it, since it acts as a watershed for

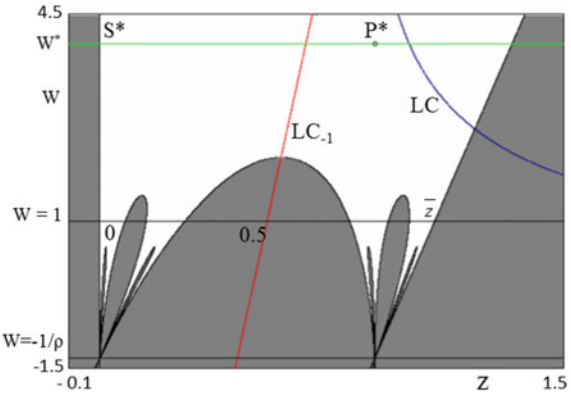

Fig. 5 The same as Fig. 4, with $\rho = 0.75$ and $\mu = 9$

the trajectories of the map T. Points belonging to $\mathcal{F}$ are mapped into $\mathcal{F}$ both under forward and backward iteration of T: more exactly $T(\mathcal{F}) \subseteq \mathcal{F}$, $T^{-1}(\mathcal{F}) = \mathcal{F}$ (see Mira et al. 1996 Chap. 5). This implies that if a saddle-point belongs to $\mathcal{F}$, then $\mathcal{F}$ must also contain the whole stable manifold (see Gumowski and Mira 1980; Mira et al. 1996). In our case, for each $\mu > 1$ and $0 \le \rho < 1$ the point $S = (0, W^*)$, located on the line of ω-limit sets $W = W^* = 1/(1-\rho)$, is a saddle point for the map T, with local stable manifold along the invariant line $z = 0$ and unstable set along the invariant line $W = W^*$. The local stable set of S belongs to $\mathcal{F}$ because the unstable manifold, along the line $W = W^*$, has a branch pointing toward the attractor P, and the opposite branch going to infinity (see Fig. 6). Then $\mathcal{F}$ includes the whole stable set of S, i.e.

$$\mathcal{F} \supseteq W^s(S) = \bigcup_{n \ge 0} T^{-n}(W^s_{loc}(S)) \tag{47}$$

where $W^s_{loc}(S)$ is given by the portion of the W axis with $W \in (0, W^*)$, denoted by ω_0 in Fig. 6, and $T^{-n}(z, W)$ denotes the set of all the rank-n preimages of the point (z, W), i.e. the set of points which are mapped into (z, W) after n applications of T. In our case, the map (44) is a noninvertible map of $Z_0 - Z_2$ type, i.e. a point (z', W') has no rank-1 preimages or two preimages, given by $T^{-1}(z', W') = T_1^{-1}(z', W') \cup T_2^{-1}(z', W')$, where

$$T_1^{-1}: \begin{cases} z = \dfrac{((W'+\mu-1)-\sqrt{(W'+\mu-1)^2-4\mu z' W'}}{2\mu} \\ W = \dfrac{W'-1}{\rho} \end{cases} \qquad T_2^{-1}: \begin{cases} z = \dfrac{((W'+\mu-1)+\sqrt{(W'+\mu-1)^2-4\mu z' W'}}{2\mu} \\ W = \dfrac{W'-1}{\rho} \end{cases} \tag{48}$$

if $\Delta(z, W) = (W' + \mu - 1)^2 - 4\mu z' W' > 0$.

We say that a point (z, W) has two preimages, given by (48), if $\Delta(z, W) > 0$, and that no inverses are defined in the points (z, W) when $\Delta(z, W) < 0$. The curve defined by the equation

$$\Delta(z, W) = (W + \mu - 1)^2 - 4\mu z W = 0, \tag{49}$$

is called critical curve LC (from the french "Ligne Critique"). Its points have two coincident preimages located on the line LC_{-1} given by

$$LC_{-1} = \{(x, y) \,|\rho W - 2\mu z + \mu = 0\} \tag{50}$$

obtained from T_i^{-1} with $\Delta = 0$. It can also be obtained as the locus of points at which the determinant of the Jacobian matrix of T vanishes, and $LC = T(LC_{-1})$ (see Gumowski and Mira 1980; Mira et al. 1996; Abraham et al. 1997). As LC_{-1} crosses the singular line δ_s out of the focal points, $LC = T\,(LC_{-1})$ is formed by two unbounded branches asymptotic to the prefocal line δ_Q (see Fig. 6). The knowledge of the curves LC and LC_{-1} is important in the computation of the preimages of the local stable set of S from which $\mathcal{F}$ is obtained according to (47). Indeed, from (48) the rank-1 preimages of $W^s_{loc}(S)$ can be easily computed. The two rank-1 preimages of ω_0, which is entirely included inside Z_2, are one on the same (invariant) W-axis and the other one on the line of equation

$$W = \frac{\mu}{\rho}(z - 1) \tag{51}$$

denoted by ω_{-1} in Fig. 6. This line intersects the line of initial conditions $W = 1$ in the point

$$\overline{z} = 1 + \frac{\rho}{\mu} \tag{52}$$

According to (47), also the line (51) belongs to $\mathcal{F}$. It can be noticed that (51) "crosses" the singular line through the focal point Q_2. The portion of this line located below the critical curve LC belongs to the region Z_2, hence it has two preimages, say ω^1_{-2} and ω^2_{-2}, whose equation can be obtained from (48) with $W' = \frac{\mu}{\rho}\left(z' - 1\right)$. These two preimages are located at opposite sides with respect to the line LC_{-1} and merge in the point H, given by the merging preimages of the point $H_1 = \omega_{-1} \cap LC$ (see Fig. 6). After some algebraic manipulation it is possible to see that such preimages belong to the curve of equation:

$$x = \frac{\mu + \rho W \pm \sqrt{(\mu + \rho W)^2 - 4\,(1 + \rho W)\left(\mu + \rho + \rho^2 W\right)}}{2\mu}. \tag{53}$$

The locus (53) represents an hyperbola if $\rho < \frac{1}{4}$, a parabola if $\rho = \frac{1}{4}$, an ellipse if $\rho > \frac{1}{4}$ (as in Fig. 6, obtained with $\rho = \frac{2}{3}$) and crosses the line with no definition

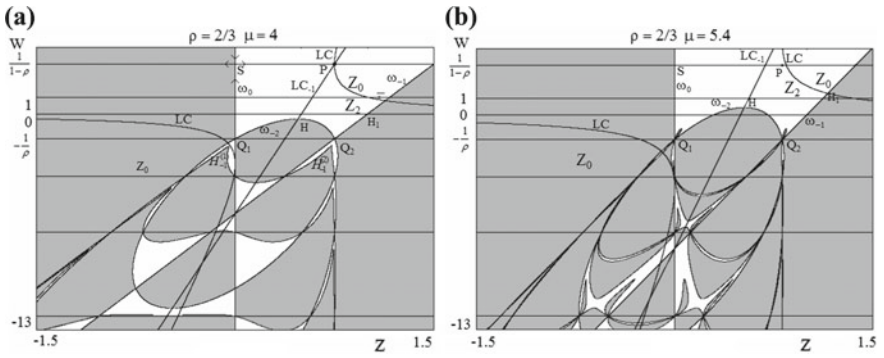

Fig. 6 An extended view of the numerically generated basins of attraction of the two-dimensional map (44): The white region represents the set of points converging to the fixed point P^* and the grey region represents the set of points which generate diverging trajectories. **a** $\rho = \frac{2}{3}$ and $\mu = 4$; **b** $\mu = 5.4$

$W = -1/\rho$ at the focal points. According to (47) also the curve (53) belongs to the frontier $\mathcal{F}$, as well as the preimages of ω_0 of any rank.

The union of all these preimages gives the boundary separating the basin $B(\infty)$ from the basin of the stable fixed point P, represented in Fig. 6, by the grey and the white regions respectively.

With the set of parameters used in Fig. 6a the two merging preimages of the point H_1, represented by the point H in which ω_{-2} intersects LC_{-1}, are below the prefocal line δ_Q. This implies that the two rank-1 preimages of H, denoted by $H_{-1}^{(1)}$ and $H_{-1}^{(2)}$ in Fig. 6a, are below the line δ_s.

As long as the point of intersection H_1 between LC and the line ω_{-1} is below the line $W = W_1 = 1$, the whole curve ω_{-2} lies below the z axis, so that the preimages of ω_{-2} are located below the singular line, as can be easily deduced from the second component of (48).

As μ increases the critical curve LC moves upwards, and when it reaches the line $W = W_1 = 1$ the curve ω_{-2} reaches the z axis, so that its preimages ω_{-3} appear, issuing from the two focal points Q_1 and Q_2. For example, in Fig. 6b the point H_1 is above the line $W = 1$, and consequently its preimage H, which is on the top of the arc ω_{-2}, is above the line $W = 0$. The two preimages of the portion of ω_{-2} above the z axis are the lobes issuing from the focal points Q_1 and Q_2, and the same happens at all the preimages of any rank of the focal points.

However, in order to understand the structure of the basins, we can limit our analysis to the portion of the plane above the line $W = -1/\rho$ (as in Fig. 7a)

For the set of parameters used in Fig. 7a the situation is similar to the one shown in Fig. 6b: the arc ω_{-2} of $\mathcal{F}$ does not intersect the line of initial conditions $W = 1$, thus it does not affect the basin of attraction $D_1(p^*)$ given by the intersection of $\mathcal{B}(\mathcal{P})$ with the line $W = 1$, according to Proposition 1. This is due to the fact that the point $H_1 = \omega_{-1} \cap LC$ is located below the line of equation $W = W_1 = 1 + \rho$, and this implies that its preimage H is located below the line of initial condition

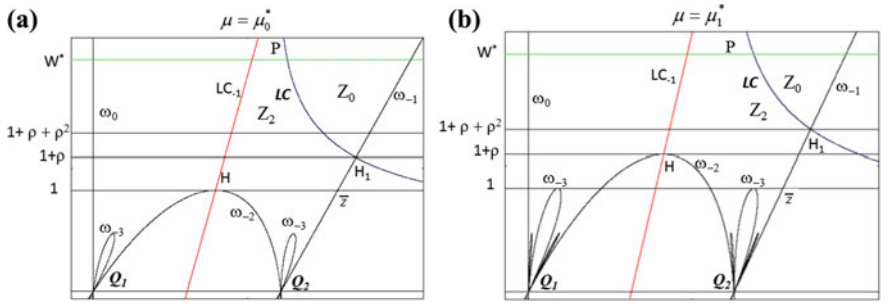

Fig. 7 Numerical computation of the preimages of the segment ω_0 located along the line $z = 0$

$W = W_0 = 1$. In fact, due to the particular structure of the second equation of the map T, the preimages of any point of a line $W = W_t$ are located on the line $W = W_{t-1}$, as can be easily computed by the second equation of (48).

As μ is increased, a value will be reached, say $\mu = \mu_0^*$, at which the point H_1 is on the line $W = W_1 = 1 + \rho$ and, as a consequence, the curve ω_{-2} becomes tangent to the line of initial conditions $W = W_0 = 1$ in the point $H = (z_H, 1)$ (see Fig. 7b) where

$$z_H = \frac{1}{2} + \frac{\rho}{2\mu}. \tag{54}$$

The value $\mu_0^* = 2 + 2\sqrt{1+\rho}$ (as can be easily computed from the tangency condition between the curve of equation (53) and the line $W = 1$) represents a bifurcation of the basin $D_1(p^*)$ of initial conditions which generate bounded price sequences. In fact for $\mu < \mu_0^*$ the basin $D_1(p^*)$ is the interval $(0, \overline{z})$, with $\overline{z}$ given by (52), whereas for $\mu > \mu_0^*$ a *hole* is created around z_H, whose points belong to $\mathcal{B}(\infty)$, bounded by the two intersections $(h_1, 1)$ and $(h_2, 1)$ between the curve (53) and the line $W = 1$.

The situation becomes even more complex as μ is further increased. The value $\mu = \mu_1^* = 2 + \rho + \sqrt{(1+\rho)\left(1+\rho+\rho^2\right)}$ is reached at which the point H_1 is on the line $W = W_2 = 1 + \rho + \rho^2$. At this value of μ two lobes of $\mathcal{B}(\infty)$, bounded by ω_{-3}, reach the line of initial conditions, the tangency points being the two preimages H_{-1}^1 and H_{-1}^2 of the point H. This gives a second bifurcation of the basin $D_1(p^*)$, at which two new holes are created around the tangency points, and the basin of the REE p^* is given by the union of 4 disjoint intervals, separated by holes of $\mathcal{B}(\infty)$.

Other similar bifurcations occur at $\mu = \mu_n^*$, where $\mu_n^* = 1 + W_n + 2\sqrt{W_n\,(1+\rho W_n)}$ at which ω_{-n-1}, belonging to the set $T^{-n-1}(\omega_0)$, become tangent to the line of initial conditions. This implies that 2^n new holes are created. The result of this sequence of bifurcations is that the basin assumes a structure which is typical of a Cantor set. In fact, at each bifurcation value $\mu = \mu_n^*$, $n \in \mathbb{N}$, the number of lobes of $B(\infty)$ is doubled, and the whole sequence of bifurcations causes a fractalization of the basin boundaries near the focal points (and their preimages) that gives a "finger-shaped" structure of $B(\infty)$. When μ reaches the limiting value

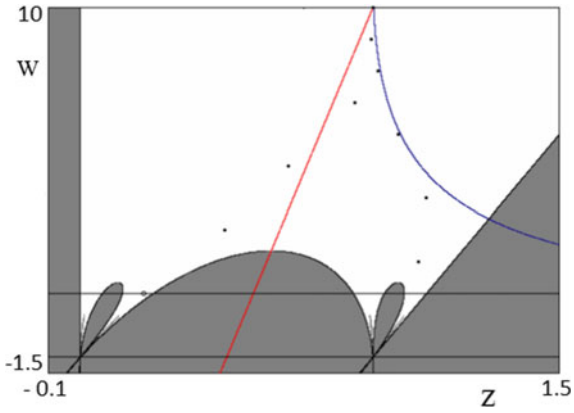

Fig. 8 The case of Bray's learnng $\rho \to 1^-$ with $\mu = 12$. The dots represent a trajectory starting from the initial condition $(z_0, 1)$ with $z_0 = 0.3$

$$\mu^*_\infty = \lim_{n\to\infty} \mu^*_n = \frac{4-\rho}{1-\rho},$$

the point H_1, together with all of its infinite preimages located at the top of the lobes, reach the line of the ω-limit sets $W = W^*$. Thus at $\mu = \mu^*_\infty$ infinitely many lobes of $B(\infty)$ have been created, and all have a contact with a chaotic attractor $\mathcal{A}$ located on the line $W = W^*$. This contact between $\partial B(\infty)$ and the chaotic set causes the disappearance of $\mathcal{A}$ and for $\mu > \mu^*_\infty$ only divergent trajectories of the map T can be obtained.

The global analysis of the basin boundaries just described holds for any value of the memory ratio ρ belonging to the interval $(0, 1)$. In particular, it also holds in the limiting case $\rho \to 1^-$. In this case the singular line, where the focal points are located, has equation $W = -1$. The equations of the curves which form the boundary $\mathcal{F}$ are obtained from those given above just substituting $\rho = 1$. So, also in the case of Bray learning (12), the complexity in the structure of the basins is conserved, as shown in Fig. 8, obtained with $\rho = 1$ and $\mu = 12$.

Similar structures of the basins are obtained for other models represented by unimodal maps, like the model proposed in Dimitri (1988), whose global analysis is given in Bischi and Naimzada (1997).

4.2 Bimodal Maps: Coexisting Stable REEs and the Problem Of equilibrium Selection

In the model analyzed above, one of the two "competing" equilibria is a rather unrealistic attractor at infinity. However, similar results hold when two or more bounded coexisting equilibria, or more complex bounded attractors of the limiting map g exist, such as periodic cycles or chaotic sets. An interesting situation arises if

a model with expectations is such that two REEs exist which are both stable under a learning rule, i.e. two "competing" rational expectation equilibria whose selection depends on the initial condition. In order to show an example where this happens, we consider a bimodal function F defined as

$$F(x) = -ax^3 + 3ax^2 - 2ax + 1 \tag{55}$$

This map has three fixed points, say $x_1^* < x_2^* < x_3^*$, as shown in Fig. 9a, where the graph of F is represented, for $a = 3$, together with the graph of $g_{0.3} = 0.7x + 0.3F(x)$, the map which governs the one-dimensional dynamics of the expected values when adaptive expectations (4) are introduced with $\alpha = 0.3$, and also represents the limiting map for the two-dimensional dynamics describing the statistical learning (18) with $\rho = 0.7$. From this graph it is evident that x_2^* is always unstable, both for F and for g, whereas for a given value of a the REEs x_1^* and x_3^* are stable for g provided that sufficiently low values of α (or sufficiently high values of ρ) are considered. So situations of two coexisting stable REEs x_1^* and x_3^* are easily obtained. In this case, the problem of equilibrium selection is related to the delimitation of the basins. Such problem is simple as far as adaptive learning (4) is concerned. In fact, the stable set of the unstable REE x_2^*, given by the set of all of its preimages, constitutes the boundary which separates the basin of x_1^* from the basin of x_3^*. These basins are formed by the two immediate basins, which include x_1^* and x_3^* respectively, and infinitely many disjoint portions, preimages of the immediate basins, which accumulate at the two periodic points of a repelling cycle $\{s_1, s_2\}$ which also constitutes the boundary of the basin of infinity, i.e. s_1 and s_2 separate the points which generate trajectories converging to bounded attractors from the ones generating unbounded trajectories.

Instead, if the two-dimensional map equivalent to the model with *statistical learning* (18) is considered, the basins appear to be more complex. In Fig. 9b the two-dimensional basins of $\left(x_1^*, W^*\right)$ and $\left(x_3^*, W^*\right)$, located along the line $W = W^* = 1/(1-\rho)$, are represented by grey and light-grey regions respectively, whereas the black region represents the basin of infinity. In this case the common boundary of the dark-grey and white regions is given by the stable set of the saddle point $\left(x_2^*, W^*\right)$ and the boundary of the basin of infinity is formed by the stable set of the saddle-cycle $\mathcal{S} = \{(s_1, W^*), (s_2, W^*)\}$. As usual, the structure of these boundaries is made up of lobes and crescents originating from the three focal points $Q_i = \left(x_i^*, -1/\rho\right)$, and the complexity of the basins of the model with statistical learning is related to the fact that the boundaries of such lobes and crescents intersect the line of initial conditions $W = 1$ in several points, so that the basins of the two stable REEs are quite different from those observed for the model with adaptive learning. For example, the trajectory starting with the initial condition $x_1^{(e)} = x_0 = 0.3 < x_2^*$, converges to x_3^*, and the one starting from $x_1^{(e)} = x_0 = 1.7 > x_2^*$, converges to x_1^*, whereas with adaptive expectations any trajectory starting with $x_0 < x_2^*$ converges to x_1^*, so that the equilibrium selection results obtained with adaptive expectations are now reversed.

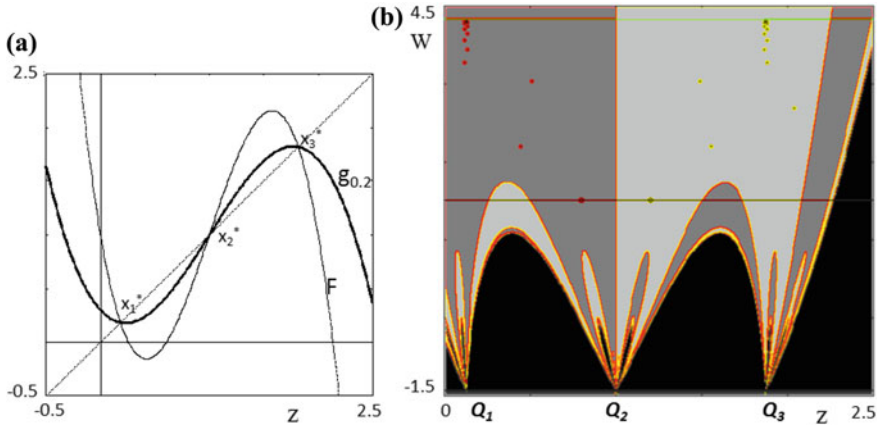

Fig. 9 **a** The graphs of the map F given in (55) for $a = 3$ together with the graph of $g_{0.3} = 0.7x + 0.3F(x)$; **b** Basins of attraction of the map (35) with F in (55): the light-grey region represents the basin of $\left(x_3^*, W^*\right)$, the grey region is the basin of $\left(x_1^*, W^*\right)$, the black region is the basin of infinity

This different equilibrium selection happens when the initial conditions are taken inside the "holes" given by the intersections of lobes and crescents with the line of initial conditions $W = 1$, whereas other initial conditions converge to the same equilibrium as in the model with adaptive expectations. This is true, for example, for the two trajectories represented in the figure, obtained with $x_1^{(e)} = 0.8$ and $x_1^{(e)} = 1.2$.

4.3 An O.G. Forward Looking Model Represented by an Increasing Map

A large class of economic problems are characterized by forward-looking expectations, i.e. are modeled by a discrete time law of motion of the form

$$x_t = F(x_{t+1}^e) \tag{56}$$

Common examples in which such mappings are obtained by Overlapping Generations (O.G.) models, where agents typically living two periods (say young and old) take the consumption and/or saving decision of their whole life in the first period (when young) so that they must guess (i.e. foresee) which will be the status of the economy (e.g. prices) one period ahead, when they will be old. As in the previous sections, x_t represents the *current (or realized) state variable* of the economic system at time t and x_{t+1}^e is the *expected state* for time $t + 1$ according to the agents' forecasting rule and their information set at time t.

Under the assumption of perfect foresight (2) the agents correctly anticipate the future state, i.e. $x_{t+1}^{(e)} = x_{t+1}$ for each t, and this defines the rational expectations

equilibrium profiles through the iteration scheme

$$x_t = F(x_{t+1}), \tag{57}$$

called *backward PF dynamics*. Indeed (57) has no dynamical meaning, but it must be simply seen as a recursive scheme which defines an *intertemporal equilibrium with PF* along which expectations are fulfilled, i.e. the equilibrium sequences generated by the recurrence (57) represent the outcomes of the economic system under the strong assumption that the agents are characterized by self-fulfilling RE.

In this case, we can have rational expectation time paths which are more complex than a stationary state. In fact, a recurrence of the form (57) can generate periodic sequences of any period or even aperiodic (i.e. chaotic) bounded sequences.

A fixed point of (57) defines a Rational Expectations Equilibrium (REE), a *k-periodic* cycle $\mathcal{C}_k = \{\alpha_1, \ldots, \alpha_k\}$, with $\alpha_i \neq \alpha_j$, $\forall\, i, j = 1, \ldots, k$, such that $\alpha_i = F(\alpha_{i+1})$, $i = 1, \ldots, k-1$, and $\alpha_k = F(\alpha_1)$, represents a Rational Expectations Cycle (REC) and so on.

In the literature on forward looking models, learning mechanisms are often proposed where the computation of x^e_{t+1} does not involve the current state x_t. Such assumption is usually motivated by saying that in modeling forward looking expectations generally the "subjects are requested to make forecasts at the beginning of period t, when x_t is not in their information set" (from Marimon et al. (1993)).

Under this assumption, the presence of x_t in both sides of the equation (56) is avoided, and it is immediate to realize that the one-dimensional dynamics which describe the time evolution of expected values of the model (56) under adaptive learning as well as the two-dimensional dynamics which describes the time evolution of expected values of the model (56) under the statistical learning, are the same as those described in the previous sections. So, the method and the results described above can be applied to many models with forward looking expectations which have been proposed in the literature. As an example, let us consider an Overlapping Generations model, proposed in Evans and Honkapohja (1995), where a representative consumer is assumed to live for two periods: period t (when young) and period $t+1$ (when old), and its utility function is $\widetilde{U} = U(c_{t+1}) - V(n_t)$, where c_{t+1} is the consumption when old, n_t the labor supply when young. In Evans and Honkapohja (1995) a concrete illustration is given, with $U(c) = \frac{c^{1-\sigma}}{1-\sigma}$, $\sigma > 0$, $V(n) = \frac{n^{1-\varepsilon}}{1-\varepsilon}$, $\varepsilon > 0$ and a production function $f(n_t, Kn_t)$, where Kn_t is the aggregate labor supply of K consumers, is considered in the separable form $f(n, Kn) = n^{\alpha}\psi(Kn)$, where $\psi(Kn) = A(I^*)^{\beta}$. These assumptions allow to obtain, for the consumer optimization problem with budget constraints

$$p^{(e)}_{t+1} c_{t+1} = M_t \quad \text{and} \quad p_t f(n_t, Kn_t) = M_t$$

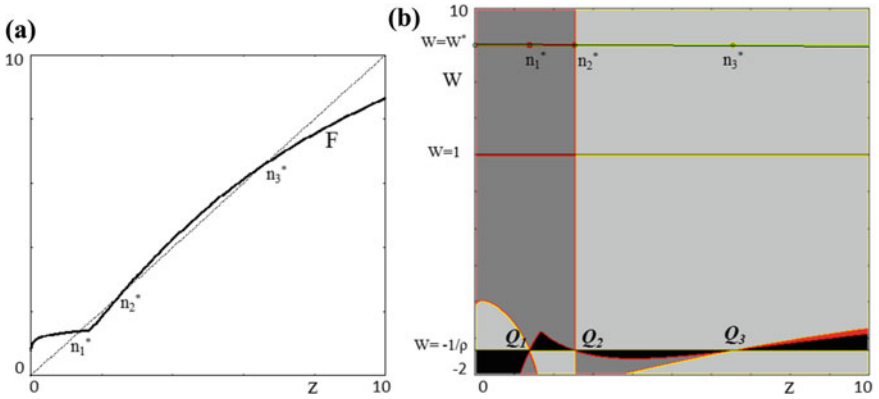

Fig. 10 **a** The graphs of the map F taken from Evans and Honkapohja (1995); **b** Basins of attraction of the map (35) with this map F: the light-grey region represents the basin of (n_3^*, W^*), the grey region is the basin of (n_1^*, W^*), the black region is the basin of infinity

a law of motion in the forward looking form

$$n_t = F(n_{t+1}^{(e)})$$

where F has a graph like the one shown in Fig. 10a, which is obtained with the same set of parameters used in Evans and Honkapohja (1995). The function F has three REEs denoted by n_i^*, $i = 1, 2, 3$, two of which are stable. In Evans and Honkapohja (1995) the following learning scheme is introduced

$$n_{t+1}^{(e)} = (1 - \alpha_t)n_t^{(e)} + \alpha_t n_{t-1} = (1 - \alpha_t)n_t^{(e)} + \alpha_t F(n_t^{(e)})$$

with $\alpha_t = 1/t$, i.e. the Bray learning (12).

If we consider the most general statistical learning (18) we obtain the same basins as the ones obtained for the limiting map $g_{1-\rho}$, i.e. bounded by the unstable REE. This is shown in Fig. 10b, where the basins of the two-dimensional map equivalent to the statistical learning (18) are represented by different grey regions. The intersection of the line of the initial conditions $W = 1$ with the intermediate-grey region represents the basin of the REE n_1^* and the intersection with the light-grey region represents the basin of the REE n_3^*. Since the portions of the stable set of the saddle point (n_2^*, W^*) which are not along the invariant line $z = n_2^*$ (i.e. the arcs originating from the focal points Q_i) are confined below the line of initial conditions $W = 1$, the basins are simply bounded by the unstable REE n_2^*, as in the case of adaptive expectations, as shown in Fig. 10b. Of course, the same holds for $\rho = 1$, i.e. in the case of Bray learning.

5 Conclusions and Further Research

The inclusion of memory of past states in discrete dynamical systems that represent economic models with expectations has been considered in the form of a weighted average with exponentially decreasing weights. This scheme is then compared to adaptive expectations. The two methods to compute expected values share the same attractors but differ for the role played by initial conditions as in general they have different basins of attraction with several coexisting attractors. So, in cases of multistability different equilibrium selections can be obtained. This result has been obtained through the study of the basins of a two-dimensional map equivalent to the statistical learning with fading memory, by using some methods for the study of global bifurcations of plane maps with a denominator that vanishes in a one-dimensional subset of the phase space. The results described in this paper have been illustrated by some simple economic examples, such as cobweb models and an overlapping generations framework. Following the path indicated by some recent works by Matsumoto and Szidarovszky in continuous-time oligopoly models with exponentially fading memory, also the methods described in this paper for discrete-time models may be usefully applied in Cournot or Bertrand oligopoly games in discrete time, see e.g. Deschamps (1975) or Thorlund-Petersen (1990). Such games, endowed with fading memory, will be reduced to an equivalent autonomous maps with denominator of dimension greater than two, a quite challenging mathematical task.

References

Abraham, R., Gardini, L., & Mira, C. (1997). *Chaos in discrete dynamical systems (A visual introduction in two dimensions)*. Springer.

Billings, L., & Curry, J. H. (1996). On noninvertible maps of the plane: Eruptions. *CHAOS*, *6*, 108–119.

Billings, L., Curry, J. H., & Phipps, E. (1997). Lyapunov exponents, singularities, and a riddling bifurcation. *Physical Review Letters*, *79*(6), 1018–1021.

Bischi, G.I., & Gardini, L. (1996) Mann iterations reducible to plane endomorphisms . In *Quaderni di Economia, Matematica e Statistica*, Facoltà di Economia (Vol. 36). Università di Urbino.

Bischi, G. I., & Naimzada, A. K. (1997). Global analysis of a nonlinear model with learning. *Economic Notes*, *26*(3), 143–174.

Bischi, G. I., & Gardini, L. (1997). Basin fractalization due to focal points in a class of triangular maps. *International Journal of Bifurcations & Chaos*, *7*(7), 1555–1577.

Bischi, G. I., Gardini, L., & Mira, C. (1999). Maps with denominator. Part 1: some generic properties. *International Journal of Bifurcation & Chaos*, *9*(1), 119–153.

Bischi, G. I., Gardini, L., & Mira, C. (2003). Plane maps with denominator. Part II: Noninvertible maps with simple focal points. *International Journal of Bifurcation & Chaos*, *13*(8), 2253–2277.

Bischi, G. I., Gardini, L., & Mira, C. (2005). Plane maps with denominator. Part III: Non simple focal points and related bifurcations. *International Journal of Bifurcation & Chaos*, *15*(2), 451–496.

Bischi, G. I., Cavalli, F., & Naimzada, A. K. (2015). Mann iteration with power means. *Journal of Difference Equations and Applications*, *21*(12), 1212–1233.

Bray, M. (1983) Convergence to rational expectations equilibrium. In R. Friedman & E. S. Phelps (Eds.), *Individual forecasting and aggregate outcomes*. Cambridge University Press.

Cavalli, F., & Naimzada, A. K. (2015). A tâtonnement process with fading memory, stabilization and optimal speed of convergence Chaos. *Solitons & Fractals*, *79*, 116–129.
Chiarella, C. (1988). The cobweb model. Its instability and the onset of chaos. *Economic Modelling*, *5*, 377–384.
Chiarella, C. (1991). The birth of limit cycles in Cournot oligopoly models with time delays. *Pure Mathematics and Applications*, *2*, 81–92.
Cushing, J. M. (1978). *Integrodifferential equations and delay models in population dynamics* (Vol. 20). Lecture notes in biomathematics. Springer.
Deschamps, R. (1975). An algorithm of game theory applied to the duopoly problem. *European Economic Review*, *6*, 187–194.
Dimitri, N. (1988) A short remark on learning of rational expectations. *Economic Notes*, 3.
Evans, G. W., & Honkapohja, S. (1995). Increasing social returns, learning and bifurcation phenomena. In A. Kirman & P. Salmon (Eds.), *Learning and rationality in economics* (pp. 216–235). Oxford: Basil Blackwell.
Foroni, I., Gardini, L., & Rosser, B, Jr. (2003). Adaptive and statistical expectations in a renewable resource market. *Mathematics and Computers in Simulation*, *63*, 541–567.
Friedman, B. M. (1979). Optimal expectations and the extreme information assumption of rational expectations macromodels. *Journal of Monetary Economics*, *5*(1), 23–41.
Fudenberg, D., & Levine, D. K. (1998) *The theory of learning in games*. The MIT Press.
Gardini, L., Bischi, G. I., & Fournier-Prunaret, D. (1999). Basin boundaries and focal points in a map coming from Bairstow's methods. *CHAOS*, *9*(2), 367–380.
Gardini, L., Bischi, G. I., & Mira, C. (2007) Maps with vanishing denominators, 16970. www.scholarpedia.org, https://doi.org/10.4249/scholarpedia.3277.
Gu, E. G., & Hao, Y.-D. (2007). On the global analysis of dynamics in a delayed regulation model with an external interference. *Chaos, Solitons & Fractals*, *34*(4), 1272–128.
Guesnerie, R., & Woodford, M. (1992) Endogenous fluctuations. In J. J. Laffont (Ed.), *Advances in economic theory* (Vol. II). Cambridge University Press.
Gumowski, I., & Mira, C. (1980). *Dynamique Chaotique*. Toulose: Cepadues editions.
Holmes, J. H., & Manning, R. (1988). Memory and market stability: The case of the cobweb. *Economic Letters*, *28*, 1–7.
Hommes, C. (1991). Adaptive learning and roads to chaos. The case of the cobweb. *Economic Letters*, *36*, 127–132.
Hommes, C. (1994). Dynamics of the cobweb model with adaptive expectations and nonlinear supply and demand. *Journal of Economic Behavior & Organization*, *24*, 315–335.
Hommes, C., Kiseleva, T., Kuznetsov, Y., & Verbic, M. (2012). Is more memory in evolutionary selection (de)stabilizing? *Macroeconomic Dynamics*, *16*, 335–357.
Hommes, C. (2013). *(2013) behavioral rationality and heterogeneous agents in complex economic systems*. Cambridge University Press.
Jensen, R. V., & Urban, R. (1984). Chaotic price dynamics in a non-linear cobweb model. *Economic Letters*, *15*, 235–240.
Lucas, R. E. (1986). Adaptive behavior and economic theory. *Journal of Business*, *59*(4).
MacDonald, N. (1978) *Time lags in biological models*. Lecture notes in biomatemathics (Vol. 27). Springer.
Marimon, R., Spear, S. E., & Sunder, S. (1993). Expectationally driven market volatility: An experimental study. *Journal of Economic Theory*, *61*, 74–103.
Marimon, R. (1997) Learning from learning in economics. In D. M. Kreps & K. F. Wallis (Eds.), *Advances in economics and econometrics: Theory and applications*, Vol. I. Cambridge University Press.
Matsumoto, A., & Szidarovszky, F. (2018) *Dynamic oligopolies with time delays*. Springer.
Matsumoto, A., & Szidarovszky, F. (2015). Dynamic monopoly with multiple continuously distributed time delays. *Mathematics and Computers in Simulation*, *108*, 99–118.
Matsumoto, A. (2017) Love affairs dynamics with one delay in losing memory or gaining affection. In A. Matsumoto (Ed.), *Optimization and dynamics with their applications*. Springer.

Mira, C., Gardini, L., Barugola, A., & Cathala, J. C. (1996). *Chaotic dynamics in two-dimensional noninvertible maps*. World Scientific.

Naimzada, A. K., & Tramontana, F. (2009). Global analysis and focal points in a model with boundedly rational consumers. *International Journal of Bifurcation & Chaos*, *19*(6), 2059–2071.

Nerlove, M. (1958). Adaptive expectations and cobweb phenomena. *Quarterly Journal of Economics*, *73*, 227–240.

Pecora, N., & Tramontana, F. (2016) Maps with vanishing denominator and their applications. *Frontiers in Applied Mathematics and Statistics*. https://doi.org/10.3389/fams.2016.00011.

Radner, R. (1983) Comment to Convergence to rational expectations equilibrium by M. Bray. In R. Friedman & E. S. Phelps (Eds.), *Individual forecasting and aggregate outcomes*. Cambridge University Press.

Thorlund-Petersen, L. (1990). Iterative computation of Cournot equilibrium. *Games and Economic Behavior*, *2*, 61–95.

Tramontana, F. (2016) Maps with vanishing denominator explained through applications in economics . *Journal of Physics: Conference Series*, *692*, Conference 1.

Yee, H. C., & Sweby, P. K. (1994). Global asymptotic behavior of iterative implicit schemes. *International Journal of Bifurcation & Chaos*, *4*(6), 1579–1611.