WORKING PAPERS IN ECONOMICS

No. 03/03

BURKHARD HEHENKAMP AND ODDVAR M. KAARBØE

IMITATORS AND OPTIMIZERS IN A CHANGING ENVIRONMENT

Department of Economics
UNIVERSITY OF BERGEN

Imitators and Optimizers in a Changing Environment

Burkhard Hehenkamp* and Oddvar M. Kaarbøe†

June 21, 2004

Abstract

We analyze the dynamic interaction between imitation and myopic optimization in an environment of changing marginal payoffs. Focusing on finite irreducible environments, we unfold a trade-off between the degree of interaction and the size of environmental shocks. The optimizer outperforms the imitator if interaction is weak or if shocks are large. We use the example of Cournot duopoly to give economic meaning to this condition. To establish our main result, we rely on continuous state space Markov theory. In particular, it turns out that introducing a stochastic environment with finitely many states suffices to make an otherwise deterministic process ergodic.

Keywords: imitation, optimization, evolution, heterogeneous learning rules, changing environments

JEL-Class.-No.: D43, D83, L13

*Department of Economics, University of Dortmund, D-44221 Dortmund, Germany. *Corresponding Author.* Email: b.hehenkamp@wiso.uni-dortmund.de. Phone: ++49-231-755-3215. Fax: ++49-231-755-3027. The paper was written during my research visit at the University of Bergen. I would very much like to thank the people at the Department of Economics for creating such a warm and inspiring research atmosphere and Ruhrgas for making this research visit possible.

†Programme for Health Economics (HEB), Department of Economics, University of Bergen. Fosswinckelsg. 6, N-5007 Bergen. Norway. Email: oddvar.kaarboe@econ.uib.no
URL http://www.econ.uib.no/stab/oddvar.html

1. Introduction

During the last decade there has emerged an exhaustive literature on the evolution of behavior in games. Only some contributions to this literature have examined the interaction of different behavioral rules. All but a few presume a stable environment in which interaction takes place. Typically, the interaction of different behavioral rules is focused on the interaction of imitative behavior and myopic optimization. Prominent examples include Conlisk (1980), Kaarbøe and Tieman (2001), Schipper (2001), and Droste et al (2002). In all these papers imitators and optimizers interact in a *stable* environment where cost and demand conditions continue to be the same every period.

As one of the main insights in this literature we have learned that the less sophisticated imitators outperform the more sophisticated (though myopic) optimizers. Imitators of successful behavior experience an evolutionary advantage in that they generally earn at least as high payoff as other types of behavior. Sometimes, and in particular in submodular (or Cournot type) games, this can lead to strictly higher payoff. In this case the imitation rule works like a commitment device, inducing the imitator to bring higher quantities to the market and leading to higher *relative* payoff. The obvious question arises when and under what circumstances a higher degree of sophistication would pay.

The purpose of our paper is to shed light on the role of stability in the environment. Therefore, we analyze the dynamic interaction between an imitator and a myopic optimizer within the changing environment of a Cournot type game. Our focus is to determine which type of behavioral rule, imitation or myopic optimization, earns higher payoff in a changing environment. To this end, we consider a model where players interact recurrently in a world of changing marginal payoff, playing a symmetric quadratic game of strategic substitutes.

In our view the assumption of a stable environment represents a strong assumption for at least two reasons. First, cost and/or demand function normally change over the business cycle. In that sense, the models referred to above, make an unrealistic assumption. More importantly, however, stability implicitly builds in an advantage for imitation. This is due to the fact that imitators always and only look into the past when deciding which quantity to put into the market. In contrast, myopic optimizers play a best response to past behavior, but take into account the current state of the environment. When demand and/or cost function can change from period to period, imitation is a more risky behavioral rule to follow than it is within a stable environment. For instance, it might be that one imitates a firm producing a high quantity last period and does not take into account that demand is lower in this period. Whenever this leads to a market price below average cost, imitators will incur a higher loss than optimizers. Similarly, if demand has increased, an optimizer will in general respond with a higher quantity, whereas the imitator will stick to some relatively low quantity. In this case, both will presumably earn positive payoff, but the optimizer a higher one. To emphasize this difference between imitation and myopic optimization, we will exclusively focus on permanently changing environments.

However, it turns out that this is not generally true. Permanently changing environments alone do not suffice to make the optimizer better off. Rather, it depends (i) on how sensitive players' actions depend on each other and (ii) on the size of environmental shocks whether the payoff advantage of imitators prevails in changing environments. Here, we unfold a trade-off between the degree of interaction and the size of environmental shocks.

Optimizers do better if the degree of interaction is sufficiently small (for some given set of environmental states). In the limit where the degree of interaction is zero this becomes palpable. In this limit, both firms will act like independent monopolists, serving separate but identical markets. On the one hand, the myopic optimizer will always choose the payoff-maximizing quantity, a monopolist would choose. This quantity will no longer depend on what the imitator did in the previous period, precisely because the degree of interaction is zero. On the other hand, the imitator will pick the quantity that the optimizer chose in the previous period, which would have been optimal in that period. Obviously, the imitator will always earn strictly less payoff than the optimizer (whenever the different environmental state implies a different monopolistic quantity). Thus, for some given environment, the optimizer will be better off, whenever interaction is sufficiently weak.

While the intuition is straightforward if interaction is weak, the picture becomes blurred when we address the reverse question. It is not at all clear whether there exists an environment for any degree of interaction such that the optimizer outperforms the imitator. To answer this question, we provide an upper boundary on the degree of interaction such that for all lower degrees the question can be answered in the positive.

Introducing changing environments, we have to decide on how to model the environmental space. We impose a natural assumption and restrict attention to irreducible environments. The main reason is that reducible environments, as the name suggests, can be decomposed into irreducible components, which then can be studied separately.

To prepare the analysis of stochastic environments, we start with investigating deterministic environments. Irreducibility implies that cost and/or demand functions follow a cyclical pattern that is independent from the starting-point. We show that, in this case, the dynamic process globally converges to a unique limit cycle and that, from some period on, the optimizer earns higher payoff every period. Contrary to most of the existing literature, such as referenced above, this represents a situation where there is selection pressure against the imitator. Therefore, this preliminary result already indicates that one critical assumption behind the advantage of imitation is a stable environment.

However, the assumption of a deterministically changing environment is comparatively strong. To give strength to our result, we proceed with analyzing the stochastic case. We show that long-run average per-period payoff is higher for the optimizer than for the imitator. Hence our claim that changes in the environment typically imply selection pressure against imitative behavior does not only hold true for cyclically changing environments, but applies to more general, stochastic environments. To establish this result, we have to impose a restriction on the set of feasible payoff parameters, capturing the above-mentioned trade-off between the degree of interaction and the size of environmental shocks. Optimizers earn higher payoff than imitators if interaction is sufficiently weak or shocks are sufficiently large.

To establish our main result, we rely heavily on the theory of Markov chains. In particular, we show that introducing a stochastic environment with finitely many states suffices to make an otherwise deterministic process ergodic. Our setup, with uncountable state space and no mutations does not allow us to apply the well-known ergodicity theorems directly, in contrast to what is typically the case in the literature (cf. e.g. Kandori *et al* 1993, Young, 1993 and Vega-Redondo, 1997). In this literature, random mutations and/or experimentation induce ergodicity of the stochastic process. As a further obstacle, our process does not satisfy the (weak) Feller property on the whole state space. It will become clear from the proof of our main theorem that this intuitively is due to the imitator's response being discontinuous on one of the zero profit lines. In consequence, we cannot apply the standard ergodicity theorems (cf. e.g. Stokey and Lucas, 1989, chap. 12).

The main step in establishing the material advantage of optimizers in changing environments is then to find a recurrent absorbing subset of the state space on which the optimizer earns strictly higher payoff than the imitator and that will be reached in an uniformly expected finite number of periods from all states outside this set. It follows that, on this subset, the imitator's response becomes a continuous function of the optimizers previous quantity. Hence, even though the process defined on the whole state space does not possess the (weak) Feller property, its restriction on the absorbing set turns out to be a T-chain – a much stronger property than the Feller property.

Notice that it is not at all clear whether any such decomposition exists. Searching for a decomposition, we encounter a number of trade-offs. For instance, if we make the absorbing set – in whatever sense – too small, then it might loose its characteristic property. On the other hand, if it is taken too large in order to make sure it is absorbing then the process might not satisfy the other desired properties, such as e.g. the (weak) Feller property. Similarly, if the set is chosen too small then the process might spend too much time in its complement set. Thus, one insight of our results below is that it is actually possible to find a decomposition, where the optimizer realizes higher payoff than the imitator on some recurrent absorbing subset of the state space.

To our knowledge, the only other papers in the (evolutionary) literature that examine the dynamic interaction between different behavioral rules in a changing environment are Gale and Rosenthal (1999), its sequel Gale and Rosenthal (2001) and Rhode and Stegeman (2001). The papers by Gale and Rosenthal study the interaction between one single experimenter and a finite number of imitators. While the experimenter randomly searches for a better strategy, the imitators adjust towards the average action of other agents. This sharply contrasts with our behavioral rule of imitation in that it is not related to success. (According to our behavioral rule of imitation, the imitator adopts the most successful action of the previous period.)

Both in our paper and in the two papers of Gale and Rosenthal, assuming a random environment makes the overall process ergodic. In contrast to Gale and Rosenthal, however, we get two important properties without relying on random experimentation. First, in terms of Gale and Rosenthal, our overall process is stable in the large. That is, it converges with probability one to the aforementioned absorbing subset of the state space. Second, it is unstable in the small: Any small subset of the absorbing set is left with probability one. The latter result is of course trivial, as it is a direct consequence of our focus on changing environments. Interestingly, a changing environment thus suffices to obtain these two properties.

Finally, Rhode and Stegeman (2001) examine a model where two players interact within an environment of changing quadratic payoff functions. In the main part of the paper, each player imitates the most successful previous action. Additionally, random noise (interpreted as imperfect control over the strategic variable) superimposes on action adjustment. In their appendix B, Rhode and Stegeman present simulations where two players with different behavioral rules – one imitator and one so-called econometrician – interact within an occasionally changing environment. The econometrician regresses the payoff function on all strategies observed in the past, relying on the correct parameter specification. Among other results they report the following (p. 451): "In particular, imitators tended to do well when structural changes were large and [random noise was] small. [...] As structural changes became larger, it was always better to be an imitator." The authors conclude that "imitation may be more profitable than apparently more sophisticated learning rules, in the presence of frequent structural change".

A first objection to Rhode and Stegeman's conclusion would of course challenge whether the adopted version of an "econometrician" is meaningful within a world of changing payoffs. Rather than elaborating on this criticism, we have chosen to stress the informational differences between imitative behavior and myopic optimization. It turns out that imitation is less profitable compared to the arguably more sophisticated learning rule of myopic optimization if interaction takes place in permanently changing environments and if structural changes are large. In that case, sophistication pays off.

The paper is organized as follows. In section 2 we outline the model, while section 3 contains the analysis. Finally, section 4 concludes.

2. The model

2.1. The stage game

We consider the following symmetric two-player game with players O and I. Every period $t \in \mathbb{N}_0$, each player chooses an action $q \in Q := [0, \overline{q}]$. Given the action profile (q^O, q^I) and payoff parameters $\pi = (\pi_{11}, \pi_{12}, \pi_1, \pi_0)$, player O's payoff is

$$u^O = -\pi_{11} \left(q^O\right)^2 /2 - \pi_{12} q^O q^I + \pi_1 q^O + \pi_0,$$

while

$$u^I = -\pi_{11} \left(q^I\right)^2 /2 - \pi_{12} q^O q^I + \pi_1 q^I + \pi_0$$

represents player I's payoff. Focussing on strategic substitutability, we assume $\pi_{11} > \pi_{12} > 0$ and $\pi_1 > 0$.

2.2. Changing environments

To model a changing environment, let payoff parameter π_1 follow a Markov process with state space $\Theta = \{\theta_1, \dots, \theta_H\}, H \geq 2$, and strictly ordered states, $0 < \theta_1 < \dots < \theta_H \leq \overline{q}$. Let $\mathcal{H} := \{1, \dots, H\}$, be the corresponding index set. If no ambuigity arises, we sometimes refer to $h \in \mathcal{H}$ as the environmental state θ_h. For simplicity, we fix the remaining environmental parameters π_{11}, π_{12}, and π_0. With slight abuse of notation, we write θ_t to denote the environmental state in period $t \in \mathbb{N}$.

The environmental process can be completely described by its transition matrix $R = (r_{ij})_{(i,j)\in\mathcal{H}^2}$. The numbers r_{ij} represent the probability of reaching environmental state $j \in \mathcal{H}$ in the next period if the current environmental state is $i \in \mathcal{H}$, i.e. $r_{ij} = \Pr\{\theta_{t+1} = \theta_j | \theta_t = \theta_i\}$. Observe that, by the Markov property, these probabilities do not depend on time. Finally, let $X := Q^2 \times \Theta$ denote the overall state space, $x = \left(q^O, q^I, \theta\right)$ represent any feasible state $x \in X$ and $\mathbf{x}_0 = \left(q_0^O, q_0^I, \theta_0\right) \in X$ be an arbitrary initial state at time $t = 0$.

As we intend to contrast our results with those obtained by other authors for the case of a stable environment, we focus on the polar opposite case. We suppose that the environment is never identical in two subsequent periods. We impose:

Assumption (CE): Let $R = (r_{ij})_{(i,j)\in\mathcal{H}^2}$ be the transition matrix of the environmental process. Then we assume $r_{ii} = 0$ for all $i \in \mathcal{H}$.

Given the environment-dependent payoff parameter π_1, both players' payoff functions depend on time via the environmental state θ_t,

$$u_t^n(x) := u_t^n(q^O, q^I, \theta_t) := -\pi_{11}\left(q^n\right)^2/2 - \pi_{12} q^O q^I + \theta_t q^n + \pi_0, \qquad \text{for } n = O, I.$$

The following example illustrates the meaning of a changing environment for the case of Cournot duopoly.

Example 1. ***(Cournot duopoly).*** *Suppose two firms engage in Cournot competition. Let $q^n \in Q$, $n = O, I$, denote the quantity produced by firm n where the cost of production q^n is $C(q^n) = c_2\left(q^n\right)^2 + c_1 q^n + c_0$. Inverse demand is $p(q^O, q^I) = a - b(q^O + q^I)$. Individual payoff is then*

$$\begin{aligned} u^n\left(q^O, q^I, \theta\right) &= p(q^O, q^I) q^n - C(q^n) \\ &= -\left(q^n\right)^2(b + c_2) + q^n(a - c_1) - b q^O q^I - c_0, \qquad \textit{for } n = O, I, \end{aligned}$$

where $\theta = a - c_1 > 0$ and $(\pi_{11}, \pi_{12}, \pi_0) = (2(b + c_2), b, -c_0)$. Accordingly, changes in the environmental state θ represent shocks in the maximum willingness to pay, a, and/or marginal cost, c_1. Since θ changes over time, the payoff function is time-dependent.

2.3. Adjustment rules

Every period t, both players update their previous action from period $t-1$. In this paper we investigate the interaction between two types of action adjustment or learning rules. While player O plays a (myopic) best response to the previous action of player I, taking into account the current state of the environment, player I chooses the action that earned highest payoff in the previous period.

Definition 1. *(Optimization) A player adjusting by myopic optimization chooses his updated action according to*

$$q_t^O := BR(q_{t-1}^I; \theta_t) := \max\{0, \frac{\theta_t}{\pi_{11}} - \rho q_{t-1}^I\}, \tag{O}$$

where $\rho := \frac{\pi_{12}}{\pi_{11}}$ *and* $\theta_t \in \Theta$.

Recall that the slope of the reaction curve, $0 < \rho < 1$, measures the degree of interaction between players. In case ρ is close to zero the degree of interaction between players is weak, and vice versa for ρ close to one.

Definition 2. *(Imitation) A player adjusting according to imitation chooses his updated action* q_t^I *such that*

$$q_t^I := \begin{cases} q_{t-1}^O & \text{if } u_{t-1}^O \geq u_{t-1}^I \\ q_{t-1}^I & \text{otherwise} \end{cases} . \tag{I}$$

Notice that we assume an imitator sticks to his strategy if and only if he has realized a strictly higher payoff. The mere purpose of this assumption is to simplify the exposition of the analysis. Alternatively, the imitator could use either action, $q_t^I \in \{q_{t-1}^O, q_{t-1}^I\}$, with some positive probability, when both players realized the same payoff in the previous period, i.e. when $u_{t-1}^O = u_{t-1}^I$. This would not affect our results.

To determine the imitator's behavior, we characterize which of the two players earns higher payoff, given some arbitrary state $x \in X$.

Lemma 1. *For any* $x = \left(q^O, q^I, \theta\right) \in X$ *we have* $u^O\left(q^O, q^I, \theta_i\right) \geq u^I\left(q^O, q^I, \theta_i\right)$ *if and only if*

$$q^O \geq q^I \text{ and } q^O + q^I \leq \frac{2\theta}{\pi_{11}}, \qquad \text{or}$$
$$q^O \leq q^I \text{ and } q^O + q^I \geq \frac{2\theta}{\pi_{11}}.$$

Proof. The claim follows from

$$\begin{aligned} u^O\left(q^O, q^I, \theta\right) - u^I\left(q^O, q^I, \theta\right) &= \frac{-\pi_{11}}{2}\left(\left(q^O\right)^2 - \left(q^I\right)^2\right) + \theta\left(q^O - q^I\right) \\ &= \left(q^O - q^I\right)\left[\frac{-\pi_{11}}{2}\left(q^O + q^I\right) + \theta\right]. \end{aligned}$$

■

Figure 1 represents the plane of the two players' quantities for any fixed environmental state $\theta \in \Theta$. With varying θ, this gives the overall state space X. Observe that q^O is depicted on the vertical axis, while q^I corresponds to the horizontal one. The figure also contains the best response curve, according to which the optimizer updates his action. The line $q^O = q^I$ indicates the response of the imitator, whenever the optimizer realizes higher payoff. Finally, the figure displays the four *relative payoff regions,* which result from the two *zero relative payoff lines* $q^O = q^I$ and $q^O + q^I = \frac{2\theta}{\pi_{11}}$ derived in Lemma 1.

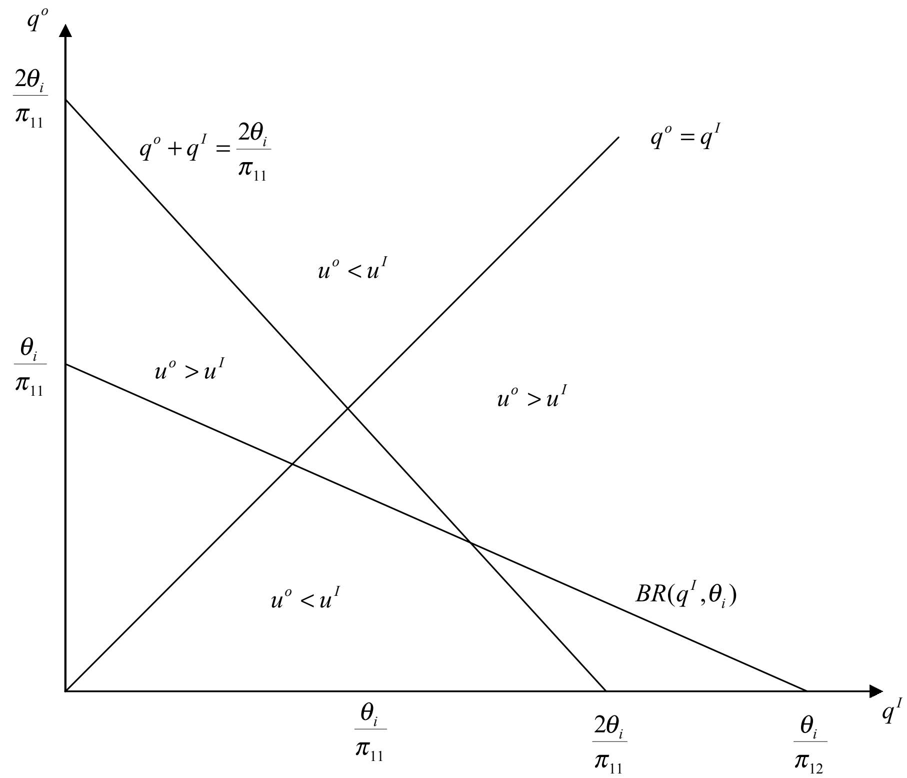

Figure 1: The action space and relative payoff

2.4. The overall process

The overall process is completely described by the environmental transition matrix R and adjustment rules (O) and (I). It induces a time-homogeneous Markov chain on the overall state space $X = Q^2 \times \Theta$, with corresponding $\sigma-$field $\mathcal{B}(X)$. The following proposition states that the overall transition function, implicitly defined by Definitions 1, 2 and the environmental Markov chain on Θ, represents a *transition probability kernel* or *Markov transition function.*

Proposition 1. *Let $P = \{P(x, A), \ x \in X, \ A \in \mathcal{B}(X)\}$ be given by (O), (I) and the Markov chain on Θ. Then*
i) for each $A \in \mathcal{B}(X)$, $P(x, A)$ is a non-negative measurable function on X, and
ii) for each $x \in X$, $P(x, \cdot)$ is a probability measure on $\mathcal{B}(X)$.

Proof. The claims can easily be established using techniques supplied in Schenk-Hoppé (1997, Prop. 1). ∎

Given the one-step transition probability kernel from Proposition 1, we can recursively define the corresponding *n-step transition probability kernal* (cp. Meyn and Tweedie, 1996, p.67). We set $P^0(x, A) := \delta_x(A)$, the Dirac measure defined by

$$\delta_x(A) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases},$$

and, for $n \geq 1$,

$$P^n(x, A) := \int_X P(x, dy) P^{n-1}(y, A),$$

where $x \in X$ and $A \in \mathcal{B}(X)$. Intuitively, $P^1(x, A) = P(x, A)$ represents the transition probability of reaching the set $A \in \mathcal{B}(X)$ when starting in state $x \in X$ in the previous period. Similarly, $P^n(x, A)$ is the probability of reaching A in precisely n periods, starting in $x \in X$ in period 0. By the Markov property we have

$$\Pr\{x_{t+n} \in A \mid x_t = x\} = P^n(x, A), \quad \forall t \geq 0.$$

Observe that the general process does not have the (weak) Feller property. This is due to the imitator's response being discontinuous on the relative payoff lines $q^O + q^I = \frac{2\theta}{\pi_{11}}, \theta \in \Theta$. As a consequence, we cannot directly apply standard decomposition results to show uniqueness of an invariant distribution. Fortunately, however, we can circumvent the problem for the important class of irreducible environments.

3. Irreducible environments

Focussing on irreducible environments, this section provides a sufficient condition on the environment such that, from some period on, the optimizer always realizes higher payoff than the imitator. This contrasts with most of the existing literature[1] in that it represents a situation displaying selection pressure against the imitator. We thus conclude that one critical assumption behind the advantage of imitation is a stable environment.

To prove this result, we develop a new decomposition technique, which – to our knowledge – has not been used in the economic literature so far. Notice that the overall process on state space X does not satisfy the (weak) Feller property. Basically this is due to the imitator's response being discontinuous on the zero relative payoff lines $q^O + q^I = \frac{2\theta_h}{\pi_{11}}$ $(h \in \mathcal{H})$. As a consequence, we cannot apply standard decomposition results to show uniqueness of an invariant distribution.

Our decomposition technique, however, allows to circumvent the problem. The basic idea is to find a decomposition of the state space into two disjoint subsets, $X = \widehat{X} \cup \widehat{X}^C$, such that the process satisfies the following two conditions. First, the complement set $\widehat{X}^C$ is uniformly transient. I.e., there exists a uniform upper boundary on the expected occupation time of $\widehat{X}^C$, i.e. applying to all $x \in \widehat{X}^C$. This uniform upper boundary then implies that the process will occupy $\widehat{X}^C$ only for an expected finite number of periods.

Second, having reached $\widehat{X}$, the process always remains in $\widehat{X}$ and the optimizer earns strictly higher payoff than the imitator. By the latter, the imitator's response is no longer discontinuous on the zero relative payoff line, entailing that the process, restricted to $\widehat{X}$, satisfies the (weak) Feller property. More specifically, we can show that the restricted process constitutes a $\varphi-$irreducible aperiodic positive Harris chain. It follows

[1]See e.g., Conlisk (1980), Rhode and Stegeman (2001), Schipper (2001), and Droste, Hommes, and Tuinstra (2002).

that the restricted process has all the desired properties such as uniqueness of an invariant distribution, the strong law of large numbers and ergodicity.

To derive $\widehat{X}$, we put a restriction on the size of shocks reflecting a trade-off between the degree of interaction – measured by the slope of the reaction function – and the size of shocks in the environment. This trade-off can be put in two ways. First, for any (finite) set of environmental states the optimizer will be better off (than the imitator) if the degree of interaction is sufficiently weak. Secondly, given some (not too high) degree of interaction, there always exists a minimum size of environmental shocks such that (for larger shocks) the optimizer outperforms the imitator.

Finally, recall that an irreducible finite environment is one where the environmental process reaches every environmental state from every other environmental state with positive probability. Since the environmental state space is finite, irreducibility implies positive recurrence, i.e. the environmental process returns to every environmental state with probability one.

We proceed as follows. In subsection 3.1, we provide the restriction and decompose the state space into two sets, $X = \widehat{X} \cup \widehat{X}^C$. Lemma 2 shows that $\widehat{X}$ is absorbing and that the optimizer earns strictly higher payoff on $\widehat{X}$. Lemma 3 establishes that $\widehat{X}^C$ is uniformly transient. To derive the stochastic properties of the restricted process, it is helpful first to analyze the case of deterministic environments, which is the core of subsection 3.2. Subsequently, subsection 3.3 examines the stochastic properties of the restricted process. In subsection 3.4, we analyze relative long-run average payoff. To this end, we construct an environmental co-chain, which allows us to derive upper and lower boundaries on long-run average payoff.

3.1. State space decomposition

The following assumption mirrors the trade-off between the degree of interaction, ρ, and the size of shocks in the environment.

Assumption (E) Let $\Theta = \{\theta_1, \dots, \theta_H\}$, $\theta_1 < \dots < \theta_H$, denote the state space and refer to the slope of the reaction function, $\rho = \frac{\pi_{12}}{\pi_{11}}$, as the degree of interaction. Then we assume that

$$\theta_{h+1} - \theta_h > 2\rho\theta_H, \qquad \text{for all } h \in \mathcal{H} \setminus \{H\}. \tag{E}$$

Notice that summing up these inequalities implies $\theta_H - \theta_1 > 2\rho(H-1)\theta_H$. Moreover, in the equidistant case $\Theta = \Theta_\theta := \{\theta, 2\theta, \dots, H\theta\}$, condition (E) becomes equivalent to $\rho < 1/(2H)$. Therefore, $H = 2$ provides an upper boundary on ρ, namely $\overline{\rho} = 1/4$, such that for any $\rho < \overline{\rho}$ there exists an environment Θ satisfying condition (E).

We start with decomposing the state space X into two subsets $\widehat{X}$ and $\widehat{X}^C$ such that $\widehat{X}$ is absorbing and $\widehat{X}^C$ is uniformly transient. First we construct $\widehat{X}$ from single pairwisely disjoint sets, $\widehat{X}_{ij}$. Part iii) of Lemma 2 below establishes that, from some period on, the process $(x_t)_{t\geq 0}$ will visit these sets $\widehat{X}_{ij}$ if and only if the previous environmental state was $\theta_{t-1} = \theta_i$ and the current environmental state is $\theta_t = \theta_j$. Recall that we focus on changing environments. Accordingly, only transitions $(i \to j)$ with $j \neq i$ are of interest, since transitions $(i \to i)$ have zero probability under the environmental process. The following index set $\mathcal{J} = \{(i,j) \in \mathcal{H}^2 : j \neq i\}$ reflects this observation.

We define $\widehat{X}_{ij}$ for $(i,j) \in \mathcal{J}$ as

$$\widehat{X}_{ij} := \left\{ x = (q^O, q^I, \theta) \in X : q^O \in [\underline{q}_j, \overline{q}_j], q^I \in [\underline{q}_i, \overline{q}_i], \theta = \theta_j \right\},$$

where $\underline{q}_h := BR(\frac{\theta_H}{\pi_{11}}; \theta_h)$ and $\overline{q}_h := BR(0; \theta_h)$ for $h \in \mathcal{H}$. To construct $\widehat{X}$ from the sets $\widehat{X}_{ij}$, we set

$$\widehat{X} := \bigcup_{(i,j) \in \mathcal{J}} \widehat{X}_{ij}.$$

Lemma 2 summarizes the properties of $\widehat{X}$.

Lemma 2. *i) The sets $\widehat{X}_{ij}$ $((i,j) \in \mathcal{J})$ form a partition of $\widehat{X}$, i.e., $\widehat{X}_{ij} \cap \widehat{X}_{i'j'} = \emptyset$, for $i \neq i'$ or $j \neq j'$ (and $\widehat{X} = \bigcup_{(i,j) \in \mathcal{J}} \widehat{X}_{ij}$, which is true by definition).*

ii) On $\widehat{X}$, the optimizer realizes higher payoff than the imitator, i.e., $u^O(x) > u^I(x)$ for all $x = (q^O, q^I, \theta) \in \widehat{X}$.

iii) $\widehat{X}$ is absorbing.

iv) If $x_T \in \widehat{X}$ for some $T < \infty$, then $x_{t+1} \in \widehat{X}_{ij} \iff (\theta_t = i$ and $\theta_{t+1} = j)$, for all $t > T$ and all $(i,j) \in \mathcal{J}$.

Proof. See Appendix. ■

Second, we decompose the complement set of $\widehat{X}$ in X, which we denote by $\widehat{X}^C := X \setminus \widehat{X}$. To this end, we define the following "level" sets:

$$\begin{aligned} X_1 \ &: = \left\{ x \in X : q^O > \frac{\theta_H}{\pi_{11}} \right\}, \\ X_2 \ &: = \left\{ x \in X : q^O \leq \frac{\theta_H}{\pi_{11}}, q^I > \frac{\theta_H}{\pi_{11}} \right\}, \\ X_3 \ &: = \left\{ x \in X : q^O, q^I \leq \frac{\theta_H}{\pi_{11}}, x \notin X_4 \cup \widehat{X} \right\}, \end{aligned}$$

where

$$X_4 := \left\{ x = (q^O, q^I, \theta) \in X : q^O \in [\underline{q}_h, \overline{q}_h],\ q^I \leq \frac{\theta_H}{\pi_{11}}, \theta = \theta_h, x \notin \widehat{X} \right\}.$$

The following Lemma shows that (i) the sets $X_1, \ldots, X_4$ form a partition of $\widehat{X}^C$; (ii) having left any level set X_i (for some level $i = 1, \ldots, 4$), the process will never return to any lower level set X_j, $j \leq i$; and (iii) the complement set $\widehat{X}^C$ is uniformly transient. The last result implies that the absorbing set $\widehat{X}$ is reached within an expected finite number of periods.

Lemma 3. *i) The four sets $X_1, ..., X_4$ form a partition of the complement set $\widehat{X}^C$.*

ii) Set $X_5 := \widehat{X}$. Then $\cup_{i=k}^5 X_i$ is absorbing for any $k = 1, \ldots, 5$.

iii) $\widehat{X}^C$ is uniformly transient.

Proof. See Appendix. ■

The proof of Lemma 3 iii) provides an upper boundary on the expected occupation time of $\widehat{X}^C$ that holds for all starting points $x_0 \in \widehat{X}^C$. Notice that restricting the set of feasible starting-points to – in whatever sense – economically reasonable ones, would allow to reduce this boundary.

Leaving $\widehat{X}^C$ in an expected finite number of periods, it is clearly warranted to investigate the properties of the process that is restricted to the absorbing set $\widehat{X}$. This is the core of the following two subsections. We start with analyzing the special case of an irreducible deterministic environment, before turning towards aperiodic environments.

3.2. Deterministic environments

The main purpose of investigating irreducible deterministic environments is to prepare the analysis of the subsequent subsection. To this end, we show that in deterministic environments behavior of the myopic optimizer and the imitator converges to a unique limit cycle. As a by-product we obtain the quantities chosen by the optimizer and the imitator along that limit cycle.

Notice that imposing deterministic transitions on the irreducible environmental process entails a cyclically changing environment. Since irreducibility implies that *all* environmental states can be reached from each other with positive probability, this cycle must be unique and visit all environmental states. Correspondingly, the length of the cycle coincides with the number of environmental states.

We first define deterministic environments and then state our results. Recall, (i) that $R = (r_{ij})_{(i,j)\in\mathcal{H}^2}$ denotes the transition matrix of the environmental (Markov) process $(\theta_t)_{t\geq 0}$, (ii) that the set of environmental states, $\Theta = \{\theta_1, \ldots, \theta_H\}$, is strictly ordered, i.e. $\theta_1 < \ldots < \theta_H$, and (iii) that $\mathcal{H} = \{1, \ldots, H\}$ represents the corresponding index set.

Definition 3. *We call the environment* deterministic *if and only if for each environmental state, $h \in \mathcal{H}$, there exists a unique successor, $s(h) := h' \in \mathcal{H}$, such that $r_{hh'} = 1$.*

Obviously, it follows from the assumption of changing environments, (CE), that $s(h) \neq h$ for all $h \in \mathcal{H}$. Moreover, as mentioned above, an irreducible, deterministic process (with finite state space) must in fact form a unique cycle of length H. The following notaton takes account of this observation.

Define the $(k+1)$*:th* successor to (the smallest) environmental state θ_1 recursively as follows: Set $s^{k+1}(\theta_1) := s(s^k(\theta_1))$ for $k = 0, \ldots, H-1$ and $s^0(\theta_1) := \theta_1$. Irreducibility implies that $(\theta_1, s^1(\theta_1) \ldots, s^{H-1}(\theta_1))$ constitutes a permutation of Θ. We then relabel the states to capture their order of appearance along the cycle, i.e. we set $\widetilde{\theta}_h := s^{h-1}(\theta_1)$

for all $h \in \mathcal{H}$ and $\widetilde{\Theta} = \{\widetilde{\theta}_1, \dots, \widetilde{\theta}_H\}$. Obviously, we have $s(\widetilde{\theta}_H) = \widetilde{\theta}_1$. Notice that we do not impose any order on subsequent states of the cycle, i.e., it may be that $\widetilde{\theta}_h > \widetilde{\theta}_{h+1}$ for some $h \in \mathcal{H} \setminus \{H\}$. Finally, we paste two cycles to the left and one the right of $\widetilde{\Theta}$, i.e., we set $\widetilde{\theta}_{h+H} := \widetilde{\theta}_{h-H} := \widetilde{\theta}_{h-2H} := \widetilde{\theta}_h$ for all $h \in \mathcal{H}$.

We can now state the main result of this subsection.

Proposition 2. *Suppose the environment is irreducible and deterministic. Then, from all starting points $x_0 = (q_0^O, q_0^I, \theta_0) \in X$, the process converges to a unique limit cycle, $(x_1^*, \dots, x_H^*)$. This cycle can be completely characterized by the optimizers quantities $(q_1^*, \dots, q_H^*)$ since, starting on the cycle, an optimizer chooses $q_t^O = q_h^*$ if and only if $\theta_t = \widetilde{\theta}_h$, while the imitator is always one period behind, playing $q_t^I = q_{t-1}^O$. Correspondingly, we can write the cycle as $x_1^* := (q_1^*, q_H^*, \widetilde{\theta}_1)$ and $x_h^* := (q_h^*, q_{h-1}^*, \widetilde{\theta}_h)$ for $h \in \mathcal{H} \setminus \{1\}$.*
If H is even, the limit cycle is given by

$$q_h^* = \frac{1}{(1 - (-\rho)^{H/2})\pi_{11}} \sum_{k=0}^{\frac{H}{2}-1} (-\rho)^k \widetilde{\theta}_{h-2k} \qquad \text{for } h \in \mathcal{H}. \tag{3.1}$$

If H is odd, we have

$$q_h^* = \frac{1}{(1 + \rho^H)\pi_{11}} \sum_{k=0}^{H-1} (-\rho)^k \widetilde{\theta}_{h-2k} \qquad \text{for } h \in \mathcal{H}. \tag{3.2}$$

Proof. See Appendix. ■

Notice that the order of actions along the limit cycle, $(q_1^*, \dots, q_H^*)$, corresponds with the order of states along the environmental cycle $(\widetilde{\theta}_1, \dots, \widetilde{\theta}_H)$.

The proof of Proposition 2 is divided into two parts. First we establish that, for deterministic environments, the property of $\widehat{X}^C$ being uniformly transient transforms into a lower boundary of $2H + 6$ periods, after which the process has left the set $\widehat{X}^C$ with certainty. The second part then deals with its behavior in $\widehat{X}$, where the process remains the rest of the time. Having reached $\widehat{X}$, the optimizer always realizes strictly higher payoff than the imitator, $u^O > u^I$, inducing the imitator to mimic the optimizer's previous action. In fact, this means that the optimizer plays a best response to his own action from two periods ago, however, taking into account the current periods' state of the environment:

$$q_{t+2}^O = BR(q_{t+1}^I; \theta_{t+2}) = BR(q_t^O; \theta_{t+2}).$$

These two-period steps imply that two cases of behavioral cycles can occur, depending on whether the number of environmental states, H, is even or odd. In the former case, we have $\min\{k \geq 1 : s^{2k}(\theta_t) = \theta_t\} = H/2$. In contrast, if H is odd, then $\min\{k \geq 1 : s^{2k}(\theta_t) = \theta_t\} = H$. Notice that both statements hold true for any $t \geq 0$.

Correspondingly, if H is even, then the limit cycle comprises two independent cycles each of length $H/2$, whereas for H being odd, there is only one cycle of length H. In the former case, the two cycles can again involve an even or odd number of states, depending on whether $H/2$ is even or odd. This explains why the feedback of the last time that quantity q_h^* was played might be positive or negative:

$$\begin{aligned} q_h^* &= BR(q_{h-2}^*; \widetilde{\theta}_h) \\ &= BR(\dots BR(q_h^*; \widetilde{\theta}_{h+2}) \dots ; \widetilde{\theta}_h) \\ &= \frac{1}{\pi_{11}} \sum_{k=0}^{\frac{H}{2}-1} (-\rho)^k \widetilde{\theta}_{h-2k} + (-\rho)^{H/2} q_h^*. \end{aligned}$$

(For notational simplicity, $q_{h-H}^* := q_{h+H}^* := q_h^*$ (for $h \in \mathcal{H}$) pastes the cycle to the left and to the right of $(q_1^*, \dots, q_H^*)$).

3.3. Aperiodic environments

Turning towards the case of aperiodic (and hence ergodic) environments, the following theorem shows that ergodicity of the environment suffices to make an otherwise deterministic process of behavioral adjustment ergodic as well.

Theorem 3.1. *Suppose the environmental process is irreducible and aperiodic. Then the restriction of P on $\widehat{X}$, $P_{\widehat{X}}$, constitutes a $\varphi-$irreducible aperiodic positive Harris chain.*

Proof. See Appendix. ■

Corollary 1. *i) The process $P_{\widehat{X}}$ admits a unique invariant distribution μ,*

$$\mu(A) = \int_{\widehat{X}} \mu(dx) P(x, A),$$

for any $A \in \mathcal{B}(\widehat{X})$.

ii) For any initial distribution λ, the distribution at time t converges to the invariant distribution μ,

$$\lim_{t\to\infty} \| \int \lambda(dx) P^t(x, \cdot) - \mu \| = 0,$$

where $\|\nu\| := \sup_{f:|f|\le 1} |v(f)| = \sup_{A\in\mathcal{B}(\widehat{X})} v(A) - \inf_{A\in\mathcal{B}(\widehat{X})} v(A)$ denotes the total variation norm.

iii) Let $\mu(g) := E_\mu[g(x_0)]$ be the steady state expectation, corresponding to the invariant measure μ. Then, the Law of Large Numbers *holds for any function g satisfying $\mu(|g|) < \infty$:*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} g(x_\tau) = \mu(g) \qquad a.s.$$

Proof. Results i) - iii) follow from Theorems 10.0.1, 13.3.1 and 17.0.1 (Meyn and Tweedie, 1996), respectively. ■

According to iii), the historical frequencies with which a certain subset, $A \in \mathcal{B}(\widehat{X})$, has been visited in the past converge to the probability mass that the invariant distribution μ assigns to A. (To see this, set $g(x_\tau) := I_A(x_\tau)$, the indicator function associated with the set $A \in \mathcal{B}(\widehat{X})$.) Similarly, *average per-period payoff* and *relative average per-period payoff* converge to their respective steady state expectations. Let $g^n(x_\tau) := u^n(x_\tau)$, $n = I, O$, and $g^{rel}(x_\tau) := u^O(x_\tau) - u^I(x_\tau)$ denote these payoffs, respectively. Then, part iii) implies $\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} g^n(x_\tau) = \mu(g^n)$, for $n = I, O, rel$. Note that we can apply the corollary, since all $g^n(\cdot)$ are bounded on $\widehat{X}$.

3.4. Long-run average payoffs

As noted earlier, Lemmas 2 and 3 characterizes the behavior of the process. In particular statement iv) in Lemma 2 implies that the process only visits those subsets $\widehat{X}_{ij}$ with $r_{ij} > 0$. We introduce the index set, $\widehat{\mathcal{J}} := \{(i,j) \in \mathcal{H}^2 : r_{ij} > 0\}$, which encompasses precisely these subsets. Because of $r_{ii} = 0$ for all $i \in \mathcal{H}$, we have that $\widehat{\mathcal{J}} \subset \mathcal{J}$. Moreover, statement iv) of Lemma 2 allows to describe the original process' transitions between these subsets $\widehat{X}_{ij}$ by a much simpler Markov co-chain with state space $\widehat{\mathcal{J}}$ and transition probabilities $R_{\widehat{\mathcal{J}}} = \left(r_{(i,j),(i',j')}\right)_{(i,j),(i',j') \in \widehat{\mathcal{J}}}$ such that

$$r_{(i,j),(i',j')} = \begin{cases} r_{jj'} & \text{if } j = i' \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 3.2 below states that the co-chain on $\widehat{\mathcal{J}}$, defined by $R_{\widehat{\mathcal{J}}}$, is irreducible and (positive) recurrent. By standard results from finite state space Markov chain theory, the co-chain then has a unique invariant distribution and the strong law of large numbers applies (cf. e.g. Resnick, 1994, sec. 2).

Theorem 3.2. *Suppose the environmental process is irreducible. Then the co-chain on $\widehat{\mathcal{J}}$ defined by $R_{\widehat{\mathcal{J}}}$ is irreducible and positive recurrent. It has a unique invariant distribution $\mu^{\widehat{\mathcal{J}}}$ and, for any bounded function $g : \widehat{\mathcal{J}} \to \mathbb{R}$ and any starting point $(i_0, j_0) \in \widehat{\mathcal{J}}$, we have*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} g(i_\tau, j_\tau) = E_{\mu^{\widehat{\mathcal{J}}}}[g], \tag{3.3}$$

where $E_{\mu^{\widehat{\mathcal{J}}}}[g] = \sum_{(i,j)\in\widehat{\mathcal{J}}} g(i,j)\mu^{\widehat{\mathcal{J}}}_{(i,j)}$.

Proof. We only establish irreducibility of the co-chain on $\widehat{\mathcal{J}}$. Positive recurrence follows from $\widehat{\mathcal{J}}$ being finite.

We have to show that for any $(i,j),(i',j') \in \widehat{\mathcal{J}}$, there exists a finite path $\widehat{\mathcal{P}} := \{(i_0,j_0),(i_1,j_1),\ldots,(i_T,j_T)\} \subset \widehat{\mathcal{J}}$ such that $(i_0,j_0) = (i,j)$, $(i_T,j_T) = (i',j')$, and $r_{(i_{t-1},j_{t-1}),(i_t,j_t)} > 0$ for all $t = 1,\ldots,T$.

Let $(i,j),(i',j') \in \widehat{\mathcal{J}}$ be arbitrary. If $j = i'$ we are done, because $(i',j') \in \widehat{\mathcal{J}}$ implies $r_{(i,j),(i',j')} = r_{jj'} = r_{i'j'} > 0$. On the other hand, if $j \neq i'$ then, by irreducibility of the environmental process, there exists a path $\mathcal{P} = \{h_0, h_1, \dots, h_T\} \subset \mathcal{H}$ such that $h_0 = j$, $h_T = i'$ and $r_{h_{t-1}h_t} > 0$ for all $t = 1, \dots, T$. Set $\widehat{\mathcal{P}} := \{(i_0, j_0), (i_1, j_1), \dots, (i_T, j_T)\}$ where $(i_0, j_0) := (i, h_0)$, $(i_t, j_t) := (h_{t-1}, h_t)$ for $t = 1, \dots, T-1$ and, finally, $(i_T, j_T) := (h_T, j')$. By construction of $\widehat{\mathcal{P}}$, we have $r_{(i_{t-1},j_{t-1}),(i_t,j_t)} = r_{j_{t-1}j_t} = r_{h_{t-1}h_t} > 0$ for all $t = 1, \dots, T$, which completes the proof. ■

If the environment is deterministic, then, by Proposition 2, the process $(x_\tau)_{\tau \in \mathbb{N}_0}$ converges to a unique limit cycle, $(x_1^*, \dots, x_H^*)$. This cycle is characterized by equations (3.1) or (3.2) and allows us to calculate long-run average per-period payoff to both the optimizer and the imitator and hence long-run *relative* average per-period payoff .

Theorem 3.3. *Suppose the environment is deterministic, let $(x_1^*, \dots, x_H^*)$ denote the unique limit cycle, and let $(x_\tau)_{\tau \in \mathbb{N}_0}$ represent any sample path of the overall process. Then we have*

$$\lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} g^{rel}(x_\tau) = \frac{1}{H} \sum_{h=1}^{H} g^{rel}(x_h^*) > 0, \tag{3.4}$$

i.e. long-run average per-period payoff is strictly higher to the optimizer than to the imitator. Moreover, from some period on, the optimizer earns strictly higher payoff than the imitator in every *period.*

Proof. The last claim follows from Proposition 2 and Lemma 2, since the process enters $\widehat{X}$ after at most $2H + 6$ periods. To show the equality, we rearrange the left hand side and get

$$\begin{aligned}
\lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} g^{rel}(x_\tau) &= \lim_{t \to \infty} \frac{1}{tH} \sum_{\tau=0}^{t-1} \sum_{h=1}^{H} g^{rel}(x_{\tau H + h}) \\
&= \frac{1}{H} \sum_{h=1}^{H} \left(\lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} g^{rel}(x_{\tau H + h}) \right) \\
&= \frac{1}{H} \sum_{h=1}^{H} g^{rel}(x_h^*),
\end{aligned}$$

where the second equality holds true, because each of the limits in the paranthese exists by Proposition 2 and Cauchy's limit theorem. Finally, the right-hand side of (3.4) is strictly positive by part ii) of Lemma 2. ■

Notice that $\frac{1}{H} \sum_{h=1}^{H} g^{rel}(x_h^*) = E_{\mu^{\widehat{\mathcal{J}}}}[g^{rel}(x_h^*)]$, since $\widehat{\mathcal{J}}$ corresponds to the limit cycle $\widetilde{\Theta} = \{\widetilde{\theta}_1, \dots, \widetilde{\theta}_H\} = \{\theta_{\kappa_1}, \dots, \theta_{\kappa_H}\}$ in that $\widehat{\mathcal{J}} = \{(\kappa_1, \kappa_2)), \dots, (\kappa_{H-1}, \kappa_H), (\kappa_H, \kappa_1)\}$ and because of $\mu^{\widehat{\mathcal{J}}} = (\frac{1}{H}, \dots, \frac{1}{H})$.

A similar result obtains for the case of aperiodic stochastic environments. In addition, Theorem 3.2 allows us to provide an upper and a lower boundary on the long-run relative average per-period payoff.

Let $g^{\max}(i,j) := \max\{g^{rel}(x) : x \in \widehat{X}_{ij}\}$ and $g^{\min}(i,j) := \min\{g^{rel}(x) : x \in \widehat{X}_{ij}\}$ denote maximum and minimum relative payoff in $\widehat{X}_{ij}$, respectively, for $(i,j) \in \widehat{\mathcal{J}}$. Since $g^{rel}(\cdot)$ is continuous and each $\widehat{X}_{ij}$ is compact, this notation is well defined. Moreover, let $\underline{g} := \min_{(i,j)\in\widehat{\mathcal{J}}} g^{\min}(i,j)$.

Theorem 3.4. *Suppose the environmental process is irreducible and aperiodic. Then long-run relative average per-period payoff converges to its steady state expectation, i.e.*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} g^{rel}(x_\tau) = E_\mu[g^{rel}(x_0)], \tag{3.5}$$

where $E_\mu[g^{rel}]$ denotes the steady state expectation, corresponding to the invariant measure μ. Moreover, the co-chain provides upper and lower boundaries on (3.5),

$$E_{\mu^{\widehat{\mathcal{J}}}}[g^{\max}] \geq E_\mu[g^{rel}(x_0)] \geq E_{\mu^{\widehat{\mathcal{J}}}}[g^{\min}] \geq \underline{g} > 0. \tag{3.6}$$

Proof. First, equality (3.5) follows from the corollary, part iii). Second, the strict inequality in (3.6) holds true by part ii) of Lemma 2. To see this, notice that $g^{rel}(x) > 0$ for all $x \in \widehat{X}$ implies

$$\underline{g} := \min_{(i,j)\in\widehat{\mathcal{J}}} g^{\min}(i,j) = \min_{(i,j)\in\widehat{\mathcal{J}}} \min\{g^{rel}(x) : x \in \widehat{X}_{ij}\} > 0,$$

since $\widehat{X}$ is compact and $g^{rel}(\cdot)$ is continuous. Third, the last weak inequality in (3.6) holds true because of $\underline{g} \leq g^{\min}(i,j)$, for all $(i,j) \in \widehat{\mathcal{J}}$.

Finally, to establish the remaining inequalities in (3.6), we make use of the co-chain $P_{\widehat{\mathcal{J}}}$. Let $\iota(x)$ denote the index function assigning the set $\widehat{X}_{ij}$ to any $x \in \widehat{X}$, i.e., $\iota(x) = ij$ if and only if $x \in \widehat{X}_{ij}$. Notice that $g^{\max}(\iota(x_t)) \geq g^{rel}(x_t) \geq g^{\min}(\iota(x_t))$, for any sample path $(x_t)_{t\geq 0}$ on $\widehat{X}$ and hence

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} g^{\max}(\iota(x_\tau)) \geq \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} g^{rel}(x_\tau) \geq \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} g^{\min}(\iota(x_\tau)). \tag{3.7}$$

By Theorem 3.2 we have that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} g^{\min}(\iota(x_\tau)) = \sum_{(i,j)\in\widehat{\mathcal{J}}} g^{\min}(i,j)\mu^{\widehat{\mathcal{J}}}_{(i,j)} = E_{\mu^{\widehat{\mathcal{J}}}}[g^{\min}]$$

and

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} g^{\max}(\iota(x_\tau)) = \sum_{(i,j)\in\widehat{\mathcal{J}}} g^{\max}(i,j)\mu^{\widehat{\mathcal{J}}}_{(i,j)} = E_{\mu^{\widehat{\mathcal{J}}}}[g^{\max}],$$

which completes the proof. ∎

To illustrate the role of Assumption (E), we reconsider the example of Cournot competition.

Example 2. ***(Cournot duopoly continued).*** *Recall that shocks were represented by changes in $\theta = a - c_1 > 0$ and that $\rho = b/(2(b + c_2))$. Notice that $\rho = 0$ (or $b = \pi_{12} = 0$) corresponds to the polar case of two independent monopolists.*
Let $\Theta = \{\theta_1, \dots, \theta_H\}$ with $0 < \theta_1 < \dots < \theta_H$ denote the environmental state space. Then Assumption (E) requires

$$\theta_{h+1} - \theta_h > 2\rho\theta_H = \frac{b}{b + c_2}\theta_H, \qquad \text{for all } h \in \mathcal{H} \setminus \{H\}. \tag{3.8}$$

As mentioned above, condition (3.8) mirrors a trade-off between the degree of interaction, ρ, on the one hand and the minimum size of shocks, $\underline{\theta} := \min\{\theta_{h+1} - \theta_h : h \in \mathcal{H} \setminus \{H\}\}$, on the other hand. In the context of Cournot duopoly, the former is related to both the degree of convexity, given by c_2, and to how responsive market price is to changes in market quantity, reflected in b.
On the one hand, for any fixed set of environmental states Θ, there exists a minimum degree of convexity $\underline{c_2}$ (given some b) such that Assumption (E) is satisfied for all $c > \underline{c_2}$ (where $\underline{c_2}$ solves $\underline{\theta}(b + \underline{c_2}) = b\theta_H$). Alternatively, given some c_2, there exists a maximum degree of price responsiveness $\overline{b}$ such that Assumption (E) applies for all $b < \overline{b}$ (let $\overline{b}$ solve $\underline{\theta}(\overline{b} + c_2) = \overline{b}\theta_H$).
On the other hand, for any fixed degree of convexity, c_2, and any fixed degree of price responsiveness, b, resulting in some $\rho < \overline{\rho}$, there exists an environment Θ and a minimum size of shocks, $b\theta_H/(b + c_2)$, such that Assumption (E) holds true for all larger shocks, i.e. if $\underline{\theta} > b\theta_H/(b + c_2)$.
In any of these cases, it follows from Theorems 3.3 and 3.4 that the optimizer earns strictly higher payoff than the imitator and hence experiences a relative evolutionary advantage.

We conclude this section by discussing the role of Assumption (E). Apparently, Assumption (E) is sufficient to establish our results in this section. To address whether is also necessary, let us check first where it enters the analysis.

First, we use Assumption (E) to show that the set X_2 will be left in $H+1$ periods with strictly positive probability. In the proof, however, we only need that $\theta_H - \theta_1 > \rho\theta_H$. This obviously represents a weaker condition than provided by assumption (E), for (E) implies $\theta_H - \theta_1 > 2\rho(H - 1)\theta_H$.

Second, Assumption (E) entails $\overline{q}_h < \underline{q}_{h+1}$ for all $h \in \mathcal{H} \setminus \{H\}$ and hence $\widehat{X}_{ij} \cap \widehat{X}_{i'j'} = \emptyset$ for $i \neq i'$ or $j \neq j'$. To guarantee these properties, we could get by on a slightly weaker condition, namely,

$$\theta_{h+1} - \theta_h > \rho\theta_H, \qquad \text{for all } h \in \mathcal{H} \setminus \{H\}. \tag{E$'$}$$

In addition, condition (E$'$) would also be sufficient for $\theta_H - \theta_1 > \rho\theta_H$.

Third and finally, we employ Assumption (E) to show that, on $\widehat{X}$, the optimizer realizes higher payoff than the imitator. In this case, Assumption (E) is *necessary* to derive that $u^O(x) > u^I(x)$ applies *for all* $x = (q^O, q^I, \theta) \in \widehat{X}$. To see this, let us drop the assumption and replace it by condition (E$'$). It follows that there exists some $h \in \mathcal{H}\backslash\{H\}$ such that $\rho\theta_H < \theta_{h+1} - \theta_h \leq 2\rho\theta_H$. Consider $\widehat{x} := (\widehat{q}^O, \widehat{q}^I, \widehat{\theta}) := (\underline{q}_h, \underline{q}_{h+1}, \theta_h)$. Obviously, we have $\widehat{x} \in \widehat{X}_{h+1,h}$ and $\widehat{q}^O \leq \overline{q}_h < \underline{q}_{h+1} \leq \widehat{q}^I$. Moreover, it follows from $\pi_{11}\underline{q}_i = \theta_i - \rho\theta_H$ that $\pi_{11}(\widehat{q}^O + \widehat{q}^I) = \theta_h + \theta_{h+1} - 2\rho\theta_H \leq 0$. Hence, by Lemma 1, we obtain $u^O(\widehat{x}) \leq u^I(\widehat{x})$.

Of course, our last observation does not mean that Assumption (E) is necessary to establish the results in this section. It only says that our technique of establishing the results requires assumption (E). However, notice that, without Assumption (E), even the restricted process will no longer display the (weak) Feller property, since the imitator's response will be discontinuous again. It will be difficult to overcome this problem.

4. Conclusions

The purpose of this paper has been to analyze the dynamic interaction between imitators and optimizers in a changing environment. To this end, we put forward a symmetric quadratic two-player game, recurrently played by one imitator and one myopic optimizer within an environment of a changing marginal payoff parameter. Restricting attention to permanently changing environments, we looked at the polar opposite case of a stable environment in order to create an as stark contrast as possible. To prepare the later analysis, we start with considering the special case of deterministically cycling environments. In these types of environments, the dynamic process globally converges to a unique limit cycle. After some finite number of periods, the process enters a subset of the state space, $\widehat{X}$, on the entire of which the optimizer earns higher payoff than the imitator. Subsequently, we investigated the more interesting case of aperiodic stochastic environments. Here, the same subset $\widehat{X}$ turns out to be the core element of the analysis. Starting from an arbitrary state of the dynamics, the process enters $\widehat{X}$ within an expected finite number of periods. We provided a uniform upper boundary on this number. Earning higher payoff on $\widehat{X}$ than the imitator, the optimizer is again better off. Thus, both scenarios represent situations in which there is selection pressure against the imitator.

Our results rely on a number of assumptions on which we will comment now. First and most importantly, we have to put a restriction on the environment capturing a trade-off between the size of environmental shocks on the one hand and the degree of interaction (=slope of the reaction function) on the other. Only if interaction is sufficiently weak or environmental shocks are sufficiently large, the optimizer will be better off. We illustrated this assumption by means of a quadratic duopoly game displaying strategic substitutes. In this context, the size of environmental shocks translates into shocks in the difference between consumers' maximum willingness to pay and a marginal cost parameter. Similarly, the degree of interaction relates to responsiveness of the market price and convexity in cost. Interaction is weak, if convexity is strong or if responsiveness is low. Correspondingly, the following circumstances will support the evolutionary advantage of optimizing behavior: highly convex cost or weakly responsive demand (implying a low degree of interaction) and/or large changes in the difference between maximum willingness to pay and marginal cost (meaning large shocks). One application fitting to our framework would be the economically important example of business cycles.

Second, focussing on irreducible environments, we only examined deterministic and aperiodic stochastic environments. While irreducibility represents a natural assumption, as we reasoned in the introduction of Section 3, it remains an open question what we can say with regard to periodic stochastic environments. As to these types of environments we encounter the problem that, in general, we do not know whether long-run payoff converges at all. Standard results for the case of countable state spaces indicate that it does so. However, even without knowing this, we have established that the absorbing set $\widehat{X}$ will be reached in an expected finite number of periods and that, on $\widehat{X}$, the optimizer earns strictly higher payoff. Thus, even without convergence of long-run payoffs, the optimizer will always be better off, once the process has reached the absorbing set.

Third, by assuming the environmental state space to be finite, we implicitly ruled out that the environmental process might be transient. In a sense, this excludes environments that represent technological progress. For, if we think that technological progress results in subsequently lower states of marginal cost and/or higher maximum willingness to pay then the environmental process might never return to former environmental states. As to this case, two scenarios obtain, depending on whether or not there exists an upper boundary on the difference between marginal cost and maximum willingness to pay. First, if the environmental state space is unbounded, then the overall process will eventually enter a subset of the state space on which the optimizer outperforms the imitator every period. Regarding average long-run payoff per period, we again face the problem of divergence. Second, if the environmental state space is bounded then shocks will be eventually too small such as to make the optimizer better off than the imitator. Under this scenario, the imitator will earn higher payoff, once the size of shocks in the environment falls below a certain threshold.

Thus, the analysis in this paper as well as our latter excursion into transient environmental spaces allow us to conclude that the advantage of imitation is strongly connected to the assumption of play taking place in a stable environment. In changing environments, optimizers do better than imitators if interaction is weak or when changes in the environment are sufficiently large. Both cases create selection pressure against imitative behavior.

References

[1] Berge, C. (1963). Topological Spaces. Oliver and Boyd, London.

[2] Conlisk J. (1980). Costly optimizers versus cheap imitators. Journal of Economic Behavior and Organization 1, 275-93.

[3] Droste, E. C. Hommes, and J. Tuinstra (2002). Endogenous flucturations under evolutionary pressure in Cournot competition. Games and Economic Behavior 40, 232-69.

[4] Gale, D. and R.W. Rosenthal (1999). Experimentation, imitation, and stochastic stability. Journal of Economic Theory 84, 1-40.

[5] Gale, D. and R.W. Rosenthal (2001). Experimentation, imitation, and stochastic stability: Addendum. Journal of Economic Theory 97, 164-174.

[6] Kaarbøe, O. and A. Tieman (2001) Evolution with mutations driven by experimentation. Mimeo, University of Bergen

[7] Kandori, M, G. Mailath, and R.Rob (1993) Learning, mutation and long run equilibria in games. Econometrica 61, 29-56.

[8] Meyn S. P. and R. L. Tweedie (1996) Markov Chains and Stochastic Stability. 3rd ed. Springer, New York, Berlin, Heidelberg.

[9] Resnick, S. (1994) Adventures in Stochastic Processes, 2nd ed., Springer, New York, Berlin, Heidelberg.

[10] Rhode Paul and Mark Stegeman (2001) Non-Nash equilibria of Darwinian dynamics with applications to duopoly. International Journal of Industrial Organization 19, 415-53.

[11] Schenk-Hoppé Klaus R. (2000) The evolution of walrasian behavior in oligopolies. Journal of Mathematical Economics 33, 35-55.

[12] Schipper Burkhard C. (2000) Imitators and optimizers in symmetric n-firm cournot oligopoly. Mimeo, University of Bonn.

[13] Stokey, N. L. and R. E. Lucas with E. C. Prescott (1989) Recursive Methods in Economic Dynamics. Harvard University Press.

[14] Tuominen, P. and R. L. Tweedie (1979) Markov chains with continuous components. Proc. London Math. Soc. (3), 38:89-114.

[15] Vega-Redondo, F. (1997). The evolution of walrasian behavior. Econometrica 65, 375-384.

[16] Young P. (1993) The evolution of conventions. Econometrica 61, 57-84.

A. Appendix

Proof of Lemma 2.

Part i): By construction of $\widehat{X}$, we only have to show that $\widehat{X}_{ij} \cap \widehat{X}_{i'j'} = \emptyset$, for $i \neq i'$ or $j \neq j'$. Recall the definition

$$\widehat{X}_{ij} := \left\{ x = (q^O, q^I, \theta) \in X : q^O \in [\underline{q}_j, \overline{q}_j], q^I \in [\underline{q}_i, \overline{q}_i], \theta = \theta_j \right\}.$$

Then the claim follows because Assumption (E) and $\theta_h < \theta_{h+1} \leq \theta_{h'}$ imply

$$\begin{aligned} \overline{q}_h &= BR(0, \theta_h) = \frac{\theta_h}{\pi_{11}} \\ &< \frac{\theta_{h+1}}{\pi_{11}} - 2\rho \frac{\theta_H}{\pi_{11}} \\ &< \frac{\theta_{h'}}{\pi_{11}} - \rho \frac{\theta_H}{\pi_{11}} = BR(\frac{\theta_H}{\pi_{11}}; \theta_{h'}) = \underline{q}_{h'}. \end{aligned}$$

Part ii): To show $u^O(x) > u^I(x)$ for all $x \in \widehat{X}$, fix $x = (q^O, q^I, \theta) \in \widehat{X}_{ij}$, for any $(i, j) \in \mathcal{J}$.

Consider $i < j$ first. On the one hand, $x \in \widehat{X}_{ij}$ implies $q^I \leq \overline{q}_i < \underline{q}_j \leq q^O$. On the other hand, it also implies $q^O + q^I \leq \overline{q}_j + \overline{q}_i < 2\theta_j/\pi_{11} = 2\theta/\pi_{11}$. The claim hence follows from Lemma 1.

Now consider $i > j$. In this case, we have $q^O \leq \overline{q}_j < \underline{q}_i \leq q^I$. Moreover, we obtain

$$\begin{aligned} q^O + q^I &\geq \underline{q}_j + \underline{q}_i \\ &= \frac{\theta_j}{\pi_{11}} - \rho \frac{\theta_H}{\pi_{11}} + \frac{\theta_i}{\pi_{11}} - \rho \frac{\theta_H}{\pi_{11}} \\ &> \frac{2\theta_j}{\pi_{11}}, \end{aligned}$$

where the last inequality holds true by Assumption (E). Again, the claim follows from Lemma 1.

Part iii): To establish that $\widehat{X}$ is absorbing, suppose $x = x_t \in \widehat{X}$. Let $s(x) = x_{t+1} \in X$ denote the direct successor to x_t under the process induced by Definitions 1 and 2. We have to show that $s(x) \in \widehat{X}$. Similar to the above definition, let $s(q^O)$ and $s(q^I)$ denote the direct successors to the optimizer's and the imitator's quantity, respectively.

First, $x \in \widehat{X}$ implies $q^O \in [\underline{q}_j, \overline{q}_j]$, $q^I \in [\underline{q}_i, \overline{q}_i]$ and $\theta = \theta_j$ for some $(i, j) \in \mathcal{J}$. Let $s(\theta_j) = \theta_k$ $(\theta_k \neq \theta_j)$ denote the subsequent environmental state that is induced by the environmental Markov chain. From Definition 1 it follows that $s(q^O) = BR(q^I; \theta_k)$. Hence $q^I \in [\underline{q}_i, \overline{q}_i] \subset [0, \theta_H/\pi_{11}]$ implies

$$s(q^O) = BR(q^I; \theta_k) \in [\underline{q}_k, \overline{q}_k]. \tag{A.1}$$

Second, by part ii) we have that $s(q^I) = q^O$ and hence $s(q^I) \in [\underline{q}_j, \overline{q}_j]$. It thus follows that $s(x) \in \widehat{X}$.

Part iv): Suppose $x_T \in \widehat{X}$ for some $T < \infty$ and let $\theta_t = \theta_i$ and $\theta_{t+1} = \theta_j$ for any $t > T$ such that $i \neq j$. Since $\widehat{X}$ is absorbing, it follows that $x_{t-1}, x_t, x_{t+1} \in \widehat{X}$ and hence $q^I_{t-1}, q^I_t \in [0, \theta_H/\pi_{11}]$. On the one hand, $q^I_t \in [0, \theta_H/\pi_{11}]$ implies $q^O_{t+1} = BR(q^I_t; \theta_j) \in [\underline{q}_j, \overline{q}_j]$. On the other hand, it follows from $\theta_t = \theta_i$ and $q^I_{t-1} \in [0, \theta_H/\pi_{11}]$ that $q^O_t = BR(q^I_{t-1}; \theta_i) \in [\underline{q}_i, \overline{q}_i]$. From part iii) we hence obtain that $q^I_{t+1} = q^O_t \in [\underline{q}_i, \overline{q}_i]$, which completes the proof of $x_{t+1} \in \widehat{X}_{ij}$.

To show the converse, fix $x_{t+1} \in \widehat{X}_{ij}$ such that $t > T$. Again, since $\widehat{X}$ is absorbing, $x_T \in \widehat{X}$ implies $x_{t-1}, x_t \in \widehat{X}$. By definition of $\widehat{X}_{ij}$, it follows that $\theta_{t+1} = \theta_j$. To show $\theta_t = \theta_i$, notice on the one hand that $x_{t-1} \in \widehat{X}$ implies $q^I_{t-1} \in [0, \theta_H/\pi_{11}]$ and hence $q^O_t = BR(q^I_{t-1}; \theta_h) \in [\underline{q}_h, \overline{q}_h]$, for some $h \neq j$. On the other hand, $x_{t+1} \in \widehat{X}_{ij}$ implies $q^I_{t+1} \in [\underline{q}_i, \overline{q}_i]$. Because of $u^O(x_t) > u^I(x_t)$, it hence follows that $q^O_t = q^I_{t+1} \in [\underline{q}_i, \overline{q}_i]$. Then however, since the intervalls $[\underline{q}_k, \overline{q}_k]$ $(k = 1, \dots, H)$ are pairwisely disjoint by Assumption (E), it must be that $[\underline{q}_h, \overline{q}_h] = [\underline{q}_i, \overline{q}_i]$, i.e. $\theta_t = \theta_i$, which completes the proof of part iv). ■

Proof of Lemma 3.

Part i): To show that the sets $X_1, \dots, X_4$ form a partition of $\widehat{X}^C$, one has to show that $\widehat{X}^C = \cup_{i=1}^4 X_i$ and $X_i \cap X_j = \emptyset$ for all $i, j = 1, \dots, 4$; $j \neq i$.

Let us establish $\widehat{X}^C = \cup_{i=1}^4 X_i$ first. Suppose $x \in X_i$ for some $i = 1, \dots, 4$. If $i = 3$ or $i = 4$ then $x \notin \widehat{X}$ holds true by definition of X_3 and X_4, respectively. For $i = 1$ and $i = 2$ the claim follows because $x \in \widehat{X}$ implies $\max\{q^O, q^I\} \leq \theta_H/\pi_{11}$, whereas $\max\{q^O, q^I\} > \theta_H/\pi_{11}$ applies for $x \in X_1 \cup X_2$. To show the opposite inclusion, fix $x \in X$ such that $x \notin X_i$ for all $i = 1, \dots, 4$. First, by definition of X_1 and X_2, we have that $x \notin X_1 \cup X_2$ implies $\max\{q^O, q^I\} \leq \theta_H/\pi_{11}$. Second, it hence follows from the definition of X_3, that $x \in X_4 \cup \widehat{X}$. Finally, $x \notin X_4$ implies $x \in \widehat{X}$. This completes the proof of $\widehat{X}^C = \cup_{i=1}^4 X_i$.

To see that $X_i \cap X_j = \emptyset$ for all $i, j = 1, \dots, 4$; $j \neq i$, notice that $x \in X_4$ implies $q^O \leq \theta_H/\pi_{11}$. Hence, $X_1 \cap X_4 = \emptyset$. The remaining intersections are empty by construction.

Part ii): Recall that $X_5 := \widehat{X}$. We have to show that $\cup_{i=k}^5 X_i$ is absorbing for any $k = 1, \dots, 5$.

First, if $k = 1$, then $\cup_{i=1}^5 X_i = X$ trivially implies $x_{t+1} \in X$, for any $x_t \in X$. Second, for $k = 2$ and arbitrary $x_t \in \cup_{i=2}^5 X_i$, we have $q^O_{t+1} = BR\left(q^I_t, \theta_{t+1}\right) \leq \frac{\theta_{t+1}}{\pi_{11}} \leq \frac{\theta_H}{\pi_{11}}$. Hence $x_{t+1} \in \cup_{i=2}^5 X_i$. Third, if $k = 3$ then $x_t \in \cup_{i=3}^5 X_i$ implies $\max\left\{q^O_t, q^I_t\right\} \leq \frac{\theta_H}{\pi_{11}}$. Therefore, it follows from $q^I_{t+1} \in \left\{q^O_t, q^I_t\right\}$ and $q^O_{t+1} \leq \frac{\theta_H}{\pi_{11}}$ that $x_{t+1} \in \cup_{i=3}^5 X_i$. Fourth, if $k = 4$ and $x_t \in X_4 \cup \widehat{X}$, we have $q^I_{t+1} \in \left\{q^O_t, q^I_t\right\} \subset [0, \theta_H/\pi_{11}]$. Let $\theta_{t+1} = \theta_h$ for some $h = 1, \dots, H$ such that $\theta_h \neq \theta_t$. Then

$$q^O_{t+1} = BR\left(q^I_t, \theta_h\right) \in \left[\underline{q}_h, \overline{q}_h\right],$$

i.e. $x_{t+1} \in X_4 \cup \widehat{X}$. Finally, for $k = 5$, the claim has been established in Lemma 2.

Part iii): We have to show that $\widehat{X}^C$ is uniformly transient, i.e., $\exists M < \infty$: $\mathbf{E}_x[\eta_{\widehat{X}^C}] \leq M$ for all $x \in X$, where $\eta_{\widehat{X}^C} := \sum_{n=1}^{\infty} I_{\{x_n \in \widehat{X}^C\}}$ denotes the number of visits of x_n to $\widehat{X}^C$. By Proposition 8.3.1 *iv*) in Meyn and Tweedie (1996, p.184), it is sufficient to show that there exists some $m \in \mathbb{Z}_+$ such that

$$P_x(\tau_{\widehat{X}^C}(m) < \infty) \leq \varepsilon < 1 \text{ for all } x \in \widehat{X}^C, \tag{A.2}$$

where $\tau_{\widehat{X}^C}(m)$ denotes the m^{th} hitting time of x_n to $\widehat{X}^C$. If this condition holds then $\mathbf{E}_x[\eta_{\widehat{X}^C}] \leq 1 + \frac{m}{1-\varepsilon}$ for all $x \in X$, that is, $\widehat{X}^C$ is uniformly transient. According to (A.2), it suffices to find a path of finite length m from $\widehat{X}^C$ to $\widehat{X}$ that has probability $1 - \varepsilon > 0$ under the overall process for all $x \in \widehat{X}^C$.

In order to derive m, we first determine, for any set X_i ($i \leq 4$), a number $m(i)$ and a probability $\varepsilon(i)$ such that $P_x(\tau_{X_i}(m(i)) < \infty) \leq \varepsilon(i) < 1$ for all $x \in X_i$. Since $\cup_{i=k}^5 X_i$ is absorbing for any $k = 1, \ldots, 5$, we can then derive m from the $m(i)$'s as $m := \sum_{i=1}^4 m(i)$. Moreover, an upper boundary $\varepsilon < 1$ satisfying (A.2) is given by $\varepsilon := 1 - (\prod_{i=1}^4 (1 - \varepsilon(i)))$.

To derive the above-mentioned numbers, we introduce the following notation. Define $s^{\tau+1}(x) := s(s^\tau(x))$, recursively, for any $\tau \geq 0$, and set $s^0(x) := x$, where $s(x)$ represents the direct successor to x under the overall process governed by the environmental process on Θ and the adjustment rules (O) and (I). Similarly, let $s^{\tau+1}(q^O)$, $s^{\tau+1}(q^I)$ and $s^{\tau+1}(\theta)$ represent the analogous operators with respect to the optimizers' quantity, the imitator's quantity and the environmental state, respectively. The operator $s^{\tau+1}(\cdot)$ gives the $(\tau + 1)$:th successor to its argument.

Case $k = 1$**:** Let $x = (q^O, q^I, \theta) \in X_1$ be arbitrary. Because of $s(q^O) = BR(q^I; s(\theta)) \leq \theta_H/\pi_{11}$, we have $s(x) \in \cup_{i=2}^5 X_i$. Since $\cup_{i=2}^5 X_i$ is absorbing, setting $m(1) := 2$ and $\varepsilon(1) := 0$ implies $P_x(\tau_{X_1}(m(1)) < \infty) \leq \varepsilon(1) < 1$ for all $x \in X_1$.

Case $k = 2$**:** Let $x = (q^O, q^I, \theta) \in X_2$ be arbitrary, i.e. $0 \leq q^O \leq \theta_H/\pi_{11} < q^I$. Since the environmental process $(\theta_t)_{t \geq 0}$ is irreducible and recurrent, there exists a finite sequence $(s^0(\theta), s^1(\theta), \ldots, s^T(\theta))$, having positive probability under the environmental process, such that $s^T(\theta) = \theta_1$ for some $0 \leq T < H$.

If there exists $\tau \in \{0, \ldots, T-1\}$ such that $u^O(s^\tau(x)) \geq u^I(s^\tau(x))$, then $s^{\tau+1}(q^I) = s^\tau(q^O)$. Since $\cup_{i=2}^5 X_i$ is absorbing, it follows that $s^\tau(q^O) \leq \theta_H/\pi_{11}$ for all $\tau > 0$. In particular, we have $s^{\tau+1}(q^I) = s^\tau(q^O) \leq \theta_H/\pi_{11}$ and $s^{\tau+1}(q^O) \leq \theta_H/\pi_{11}$ and hence $s^{\tau+1}(x) \in \cup_{i=3}^5 X_i$.

If, to the contrary, $u^O(s^\tau(x)) < u^I(s^\tau(x))$ for all $\tau = 0, \ldots, T-1$, then $s^\tau(q^I) = q^I > \theta_H/\pi_{11}$ for all $\tau = 1, \ldots, T$. In particular, this implies

$$s^T(q^O) \leq \theta_H/\pi_{11} < s^T(q^I). \tag{A.3}$$

Moreover, it follows that

$$\begin{aligned} s^T(q^O) + s^T(q^I) &= \frac{\theta_1}{\pi_{11}} - \rho s^{T-1}(q^I) + s^{T-1}(q^I) \\ &= \frac{\theta_1}{\pi_{11}} + (1-\rho) q^I \\ &> \frac{\theta_1}{\pi_{11}} + (1-\rho)\frac{\theta_H}{\pi_{11}} > \frac{2\theta_1}{\pi_{11}}, \end{aligned} \tag{A.4}$$

where the last inequality follows from Assumption (E) because of

$$\theta_1 < (1 - 2\rho(H-1))\theta_H < (1-\rho)\theta_H$$

which, in turn, holds true because adding up the inequalities in Assumption (E) yields

$$\theta_H - \theta_1 > 2(H-1)\rho\theta_H. \tag{A.5}$$

By Lemma 1, inequalities (A.3) and (A.4) imply $u^O(s^T(x)) > u^I(s^T(x))$ so that $s^{T+1}(x) \in \cup_{i=3}^5 X_i$.

Thus, in both cases we have $s^{T+1}(x) \in \cup_{i=3}^5 X_i$ for some $T < H$.

We can now derive the values for $\varepsilon(2)$ and $m(2)$. To define $\varepsilon(2)$, let r_{h1} be the joint probability of all environmental paths from θ_h to θ_1 ($h > 1$) that involve strictly less than H transitions. Similarly, let r_1 be the minimum of all joint probabilities r_{h1}, i.e. $r_1 := \min_{h>1} r_{h1}$. Since the environmental process is irreducible it must be that $r_{h1} > 0$, for all $h > 1$, and hence $r_1 > 0$. Thus, setting $m(2) := H + 1$ and $\varepsilon(2) := 1 - r_1$ implies $P_x(\tau_{X_2}(m(2)) < \infty) \leq \varepsilon(2) < 1$ for all $x \in X_2$.

Case $k = 3$**:** Let $x = (q^O, q^I, \theta) \in X_3$ be arbitrary so that $0 \leq q^O, q^I \leq \theta_H/\pi_{11}$. It follows that $s(q^O) = BR(q^I, s(\theta)) \in [\underline{q}_h, \overline{q}_h]$ for some $h = 1, \dots, H$ such that $s(\theta) = \theta_h$. Thus, $s(x) \in \cup_{i=4}^5 X_i$ so that $m(3) := 2$ and $\varepsilon(1) := 0$ imply $P_x(\tau_{X_1}(m(3)) < \infty) \leq \varepsilon(3) < 1$ for all $x \in X_3$.

Case $k = 4$**:** Let $x = (q^O, q^I, \theta) \in X_4$ be arbitrary, i.e. $0 \leq q^O, q^I \leq \theta_H/\pi_{11}$ and $q^O \in [\underline{q}_h, \overline{q}_h]$ such that $\theta_h = \theta$. Define $\widetilde{q}$ implicitly by

$$BR(\widetilde{q}; \theta_1) + \widetilde{q} = \frac{2\theta_1}{\pi_{11}}. \tag{A.6}$$

We distinguish two cases depending on whether $q^I > \widetilde{q}$ or not. Notice that (A.6) is equivalent to

$$(1-\rho)\widetilde{q} = \frac{\theta_1}{\pi_{11}}. \tag{A.7}$$

Consider $q^I > \widetilde{q}$ first. Similar to case $k = 2$ above, there exists a finite sequence of environmental states, $(s^0(\theta), s^1(\theta), \dots, s^T(\theta))$, having positive probability under the environmental process, such that $s^T(\theta) = \theta_1$ for some $T < H$. On the one hand, if $u^O(s^\tau(x)) \geq u^I(s^\tau(x))$ for some $\tau = 0, \dots, T-1$, then $s^{\tau+1}(q^I) = s^\tau(q^O) \in [\underline{q}_i, \overline{q}_i]$, where i satisfies $\theta_i = s^\tau(\theta)$. Similarly, $s^{\tau+1}(q^O) = BR(s^\tau(q^I); \theta_j) \in [\underline{q}_j, \overline{q}_j]$, such that $s^{\tau+1}(\theta) = \theta_j$. It follows that $s^{\tau+1}(x) \in \widehat{X}_{ij}$, which implies $s^{\tau+1}(x) \in X_5$ because of $\widehat{X}_{ij} \subset \widehat{X} = X_5$. On the other hand, if $u^O(s^\tau(x)) < u^I(s^\tau(x))$ for all $\tau = 0, \dots, T-1$, then $s^\tau(q^I) = q^I$ for all $\tau = 1, \dots, T$. It follows that $s^T(q^I) = q^I > \widetilde{q}$ and

$$s^T(q^O) = BR(s^{T-1}(q^I); s^T(\theta)) = BR(q^I; \theta_1) = \frac{\theta_1}{\pi_{11}} - \rho q^I.$$

From (A.7), we hence obtain

$$
\begin{aligned}
s^T(q^O) + s^T(q^I) &= \frac{\theta_1}{\pi_{11}} + (1-\rho)q^I \\
&> \frac{\theta_1}{\pi_{11}} + (1-\rho)\widetilde{q} = \frac{2\theta_1}{\pi_{11}}.
\end{aligned} \tag{A.8}
$$

Moreover, $\widetilde{q} < q^I$ implies $\theta_1/\pi_{11} < (1-\rho)q^I < (1+\rho)q^I$ and hence

$$
s^T(q^O) = \frac{\theta_1}{\pi_{11}} - \rho q^I < q^I = s^T(q^I).
$$

By Lemma 1, we thus have $u^O(s^T(x)) > u^I(s^T(x))$. An argument similar to the one above (now applied w.r.t. T rather than τ) shows $s^{T+1}(x) \in \widehat{X}_{ij} \subset X_5$, where $\theta_i = s^T(\theta)$ and $\theta_j = s^{T+1}(\theta)$.

Second, consider $q^I \leq \widetilde{q}$. Again, since the environmental process is irreducible, there exists a finite sequence $(s^0(\theta), s^1(\theta), \ldots, s^T(\theta))$, having positive probability, such that $s^T(\theta) = \theta_H$ for some $T < H$. If $u^O(s^\tau(x)) \geq u^I(s^\tau(x))$ for some $\tau = 0, \ldots, T-1$, then, by the same argument used above, we have $s^{\tau+1}(x) \in \widehat{X}_{ij} \subset X_5$, where $\theta_i = s^\tau(\theta)$ and $\theta_j = s^{\tau+1}(\theta)$. On the other hand, if $u^O(s^\tau(x)) < u^I(s^\tau(x))$ for all $\tau = 0, \ldots, T-1$, then $s^\tau(q^I) = q^I$ for all $\tau = 1, \ldots, T$. In particular, we obtain $s^T(q^I) = q^I \leq \widetilde{q}$ and

$$
s^T(q^O) = BR(s^{T-1}(q^I); s^T(\theta)) = BR(q^I; \theta_H) = \frac{\theta_H}{\pi_{11}} - \rho q^I.
$$

First, from (A.7), it follows that

$$
\begin{aligned}
s^T(q^O) + s^T(q^I) &= \frac{\theta_H}{\pi_{11}} + (1-\rho)q^I \\
&\leq \frac{\theta_H}{\pi_{11}} + (1-\rho)\widetilde{q} \\
&= \frac{\theta_H}{\pi_{11}} + \frac{\theta_1}{\pi_{11}} < \frac{2\theta_H}{\pi_{11}}.
\end{aligned} \tag{A.9}
$$

Second, notice that $s^T(q^O) > s^T(q^I)$ is equivalent to $\theta_H/\pi_{11} > (1+\rho)q^I$. Therefore, it is sufficient to show that

$$
(1+\rho)\theta_1 < (1-\rho)\theta_H, \tag{A.10}
$$

since (A.10) and (A.7) imply

$$
(1+\rho)q^I \leq (1+\rho)\widetilde{q} = \frac{\theta_1(1+\rho)}{\pi_{11}(1-\rho)} < \frac{\theta_H}{\pi_{11}}.
$$

To see (A.10), notice that (A.5) is equivalent to

$$
\theta_H(1-\rho(H-1)) > \theta_1 + \rho(H-1)\theta_H,
$$

and hence implies

$$\theta_H(1-\rho) > \theta_H(1-\rho(H-1)) > \theta_1 + \rho(H-1)\theta_H > \theta_1(1+\rho),$$

which yields $s^T(q^O) > s^T(q^I)$. Combined with (A.9), it thus follows from Lemma 1 that $u^O(s^T(x)) > u^I(s^T(x))$. The same argument as the one applied for the case $q^I > \widetilde{q}$ then shows $s^{T+1}(x) \in \widehat{X}_{ij} \subset X_5$, where $\theta_i = s^T(\theta)$ and $\theta_j = s^{T+1}(\theta)$.

We can now derive the values for $\varepsilon(4)$ and $m(4)$, respectively. To define $\varepsilon(4)$, recall (i) that r_{h1} denotes the joint probability of all environmental paths from θ_h to θ_1 $(h > 1)$ involving strictly less than H transitions, (ii) that r_1 represents the minimum of all these joint probabilities r_{h1}, i.e. $r_1 := \min_{h>1} r_{h1}$, and (iii) that $r_1 > 0$. Similarly, (i) let r_{hH} be the joint probability of all environmental paths from θ_h to θ_H $(h < H)$ involving strictly less than H transitions; (ii) let r_H be the minimum of all joint probabilities r_{hH}, i.e. $r_H := \min_{h>1} r_{hH}$ and notice (iii) that it must be $r_{hH} > 0$, for all $h < H$, and hence $r_H > 0$, since the environmental process is irreducible. Fix $r := \min\{r_1, r_H\} > 0$.

Then, setting $m(4) := H+1$ and $\varepsilon(4) := 1-r$ implies $P_x(\tau_{X_2}(m(4)) < \infty) \le \varepsilon(4) < 1$ for all $x \in X_4$, which completes the proof of Lemma 3. ■

Proof of Proposition 2. Suppose the environment is irreducible and deterministic. The proof is divided into two parts. We first establish that $\widehat{X}$ is reached within $2H+6$ periods. Subsequently, we show that, having entered $\widehat{X}$, the process converges to a unique limit cycle.

Part i): To show that $\widehat{X}$ is reached within $2H+6$ periods, recall the numbers $m(k)$, $k = 1, \ldots, 4$, provided in the proof of Lemma 3 iii). Notice that, for deterministic environments, these numbers translate into lower boundaries after which the respective level set X_k has been left *with certainty.* Correspondingly, we can set $\varepsilon(k) := 0$, for all $k = 1, \ldots, 4$. It follows that $m = \sum_{k=1}^4 m(k) = 2H+6$ and $\varepsilon = 1 - \prod_{k=1}^4 (1-\varepsilon(k)) = 0$, which finally implies

$$P_x(\tau_{\widehat{X}^C}(2H+6) < \infty) = 0, \quad \text{for all } x \in \widehat{X}^C.$$

Thus, $\widehat{X}$ is reached within $2H+6$ periods.

Part ii): Fix $T < \infty$ such that $x_T \in \widehat{X}$. By Lemma 2, it follows that $u^O(x_t) > u^I(x_t)$ and hence $q^I_{t+1} = q^O_t$ for all $t \ge T$. Hence,

$$q^O_{t+2} = BR(q^I_{t+1}; \theta_{t+2}) = BR(q^O_t; \theta_{t+2}) \quad \text{for all } t \ge T. \tag{A.11}$$

Iterative application of (A.11) yields

$$\begin{aligned} q^O_{T+2K} &= BR(q^O_{T+2K-2}; \theta_{T+2K}) \\ &= BR(\ldots BR(q^O_T; \theta_{T+2}) \ldots ; \theta_{T+2K}) \\ &= \frac{\theta_{T+2K}}{\pi_{11}} - \rho\left[\frac{\theta_{T+2K-2}}{\pi_{11}} - \rho\left[\ldots\left[\frac{\theta_{T+2}}{\pi_{11}} - \rho q^O_t\right]\right]\right] \\ &= \frac{1}{\pi_{11}} \sum_{k=0}^{K-1} (-\rho)^k \theta_{T+2(K-k)} + (-\rho)^K q^O_T, \end{aligned} \tag{A.12}$$

for any $K \in \mathbb{N}$. For some arbitrary $\theta_t \in \Theta$, let $\overline{K} := \min\{k \in \mathbb{N} : \theta_{t+2k} = \theta_t\}$ denote the minimum number of double periods such that the environmental process returns to θ_t. Notice first that $\overline{K}$ does not depend on θ_t, since the environment is deterministic. Second, we have $\overline{K} = H$ if H is odd, whereas $\overline{K} = H/2$ holds true for H being even.

Next, we make use of the cycling environment, which implies that $\theta_t = \theta_{t+2l\overline{K}}$ for all $l \in \mathbb{N}$ and all $t \geq T$. Let $L \in \mathbb{N}$. Then, substituting $K := L\overline{K}$ in equation (A.12) yields

$$\begin{aligned} q^O_{T+2L\overline{K}} &= \frac{1}{\pi_{11}} \sum_{k=0}^{L\overline{K}-1} (-\rho)^k \theta_{T+2(L\overline{K}-k)} + (-\rho)^{L\overline{K}} q^O_T \\ &= \frac{1}{\pi_{11}} \sum_{l=0}^{L-1} \sum_{k=0}^{\overline{K}-1} (-\rho)^{l\overline{K}+k} \theta_{T+2(L\overline{K}-(l\overline{K}+k))} + (-\rho)^{L\overline{K}} q^O_T \\ &= \frac{1}{\pi_{11}} \left(\sum_{l=0}^{L-1} (-\rho)^{l\overline{K}} \left(\sum_{k=0}^{\overline{K}-1} (-\rho)^k \theta_{T+2(L-l)\overline{K}-2k} \right) \right) + (-\rho)^{L\overline{K}} q^O_T \\ &= \frac{1}{\pi_{11}} \left(\sum_{l=0}^{L-1} (-\rho)^{l\overline{K}} \right) \left(\sum_{k=0}^{\overline{K}-1} (-\rho)^k \theta_{T+2\overline{K}-2k} \right) + (-\rho)^{L\overline{K}} q^O_T. \end{aligned} \tag{A.13}$$

Because of $\rho < 1$, we can take the limit $L \to \infty$ in (A.13) to obtain

$$\lim_{L\to\infty} q^O_{T+2L\overline{K}} = \frac{1}{(1-(-\rho)^{\overline{K}})\pi_{11}} \sum_{k=0}^{\overline{K}-1} (-\rho)^k \theta_{T+2\overline{K}-2k}. \tag{A.14}$$

If H is odd, we have $\overline{K} = H$ and hence (A.14) results in

$$\lim_{L\to\infty} q^O_{T+2L\overline{K}} = \frac{1}{(1+\rho^H)\pi_{11}} \sum_{k=0}^{H-1} (-\rho)^k \theta_{T+2H-2k},$$

whereas for H being even, $\overline{K} = H/2$ implies

$$\lim_{L\to\infty} q^O_{T+LH} = \frac{1}{(1-(-\rho)^{H/2})\pi_{11}} \sum_{k=0}^{\frac{H}{2}-1} (-\rho)^k \theta_{T+H-2k}.$$

The claim then holds true, since we can choose $T \geq \min\{t : x_t \in \widehat{X}\}$ and hence $\theta_T = \widetilde{\theta}_h$ and $h \in \mathcal{H}$ arbitrarily (where $h \in \mathcal{H}$ denotes the position of θ_T in the environmental cycle $(\widetilde{\theta}_1, \ldots, \widetilde{\theta}_H)$). ■

Proof of Theorem 3.1. The proof is divided into the following parts.

i) $P_{\widehat{X}}$ is well-defined, i.e., $P_{\widehat{X}}(x, A) := P\left(x, A \cap \widehat{X}\right)$ defines a transition probability kernel;

ii) $P_{\widehat{X}}$ is a (weak) Feller chain;

iii) $P_{\widehat{X}}$ is a T-chain;

iv) $P_{\widehat{X}}$ is Harris recurrent;

v) $P_{\widehat{X}}$ is aperiodic;

vi) $P_{\widehat{X}}$ is $\varphi-$irreducible;

vii) $P_{\widehat{X}}$ is positive.

Part i): We will show that $P_{\widehat{X}}(x, A) := P(x, A \cap \widehat{X})$ defines a transition probability kernel (for a definition, see Meyn and Tweedie, 1996, p. 65). First, for all $A \in \mathcal{B}(\widehat{X})$, $P_{\widehat{X}}(\cdot, A)$ is a non-negative measurable function on $\widehat{X}$, because $P(\cdot, A \cap \widehat{X})$ is a non-negative measurable function on X, for all $A \in \mathcal{B}(\widehat{X})$, which follows from Proposition 1. Second, for all $x \in \widehat{X}$, $P_{\widehat{X}}(x, \cdot)$ is a probability measure on $\mathcal{B}(\widehat{X})$, because $\widehat{X}$ is absorbing. This implies

$$P(x, A) = P\left(x, A \cap \widehat{X}\right) + \underbrace{P\left(x, A \cap (X \setminus \widehat{X})\right)}_{=0} = P_{\widehat{X}}(\cdot, A),$$

for all $x \in \widehat{X}$, which completes the proof of Part i).

Part ii): By definition, a transition probability kernel is a (weak) Feller chain if and only if $P(\cdot, O)$ is lower semicontinuous for any open set $O \in \mathcal{B}(X)$ (cf. Meyn and Tweedie, 1996, Sec. 6, p. 127). Therefore, it is sufficient to show that the level sets $\{x \in \widehat{X} : P_{\widehat{X}}(x, \mathcal{O}) \leq c\}$ are topologically closed, for any $c \in \mathbb{R}$ and any open set $\mathcal{O} \in \mathcal{B}(\widehat{X})$ (cf. Meyn and Tweedie, 1996, Appendix D.4, p. 520).

Let $\mathcal{O} \in \mathcal{B}(\widehat{X})$ be such an open set and fix $c \in \mathbb{R}$. Observe

$$\left\{x \in \widehat{X} : P(x, \mathcal{O}) \leq c\right\} = \bigcup_{k \in \mathcal{H}} \left\{x \in \widehat{X}_h : P(x, \mathcal{O}) \leq c\right\},$$

where $\widehat{X}_k := \cup_{g \in \mathcal{H} \setminus \{k\}} \widehat{X}_{gk}$, for $k \in \mathcal{H}$, and

$$\begin{aligned}\left\{x \in \widehat{X}_k : P(x, \mathcal{O}) \leq c\right\} &= \left\{x \in \widehat{X}_k : P\left(x, \cup_{(i,j) \in \mathcal{J}} (\mathcal{O}_{ij})\right) \leq c\right\} \\ &= \left\{x \in \widehat{X}_k : \sum_{(i,j) \in \mathcal{J}} P(x, \mathcal{O}_{ij}) \leq c\right\},\end{aligned} \tag{A.15}$$

where $\mathcal{O}_{ij} := \mathcal{O} \cap \widehat{X}_{ij}$. Consequently, it is sufficient to show that the last set in (A.15) is closed, for any $k \in \mathcal{H}$. Moreover, since the sum of lower semicontinuous functions constitutes a lower semicontinuous function itself (cf. Berge, 1963, p. 77, Theorem 5), we are done if we can show that the sets $\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) \leq c\}$ are closed for any set

$\mathcal{O}_{ij} \in \mathcal{B}(\widehat{X}_{ij})$ that is open relative to $\widehat{X}_{ij}$. Notice that, since the sets $\widehat{X}_{ij}$, $(i,j) \in \mathcal{J}$, are disconnected, $\mathcal{O}$ is open relative to $\widehat{X}$ if and only if each $\mathcal{O}_{ij}$ is open relative to $\widehat{X}_{ij}$ (i.e. for all $(i,j) \in \mathcal{J}$).

Let $(i,j) \in \mathcal{J}$ be arbitrary and let $\mathcal{O}_{ij} \in \mathcal{B}(\widehat{X}_{ij})$ be open relative to $\widehat{X}_{ij}$. Observe first that $P(x, \mathcal{O}_{ij}) \in \{0, r_{ij}\}$ for all $x \in \widehat{X}$ and all $(i,j) \in \mathcal{J}$ and that, second, $P(x, \mathcal{O}_{ij}) > 0$ only if $x \in \widehat{X}_i$.

If $c < 0$ then $\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) \leq c\} = \emptyset$, which is closed. If $c \in [0, r_{ij})$ then

$$\begin{aligned}\left\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) \leq c\right\} &= \left\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) = 0\right\} \\ &= \left\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) > 0\right\}^C .\end{aligned}$$

We show that $\left\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) > 0\right\}$ is open. To this end, define $f_{ij} : \widehat{X}_i \mid_{Q^2} \rightarrow \widehat{X}_j \mid_{Q^2}$ for $(i,j) \in \mathcal{J}$ by

$$f_{ij}\left(q^O, q^I\right) := \left(BR\left(q^I; \theta_j\right), q^O\right).$$

The f_{ij}'s represent the deterministic transitions on the action space Q^2 given the environment makes a transition from θ_i to θ_j and taking into account that $u^O > u^I$ on $\widehat{X}$. Observe that each f_{ij} is continuous. Furthermore, for any $k = L, H$,

$$\left\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) > 0\right\} = \left\{x = \left(q^O, q^I, \theta_k\right) \in \widehat{X}_k : \exists y \in \mathcal{O}_{ij} \text{ s.t. } y = f_{ij}\left(q^O, q^I\right)\right\}.$$

On the one hand, if $k \neq i$ then this set is empty and hence open. On the other hand, if $k = i$ then this set is open because it represents the inverse image of the open set $\mathcal{O}_{ij}$ under the continuous function f_{ij}. Thus, $\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) \leq c\}$ is closed, for any $c \in [0, r_{ij})$.

Finally, if $c \geq r_{ij}$ then $\{x \in \widehat{X}_k : P(x, \mathcal{O}_{ij}) \leq c\} = \widehat{X}_k$, which is closed.

Part iii): $P_{\widehat{X}}$ is a $T-$chain.

We have to establish that there is a distribution $a = \{a(t)\}_{t=0}^{\infty}$ on $\mathbb{N}$ and a sub-stochastic transition kernel T such that

$$a) \quad K_a(x, A) := \sum_{t=0}^{\infty} P^t(x, A)\, a(t) \geq T(x, A) \qquad \forall x \in \widehat{X}, \forall A \in \mathcal{B}(\widehat{X}),$$

$$b) \quad T(\cdot, A) \text{ is a lower semicontinuous function } \forall A \in \mathcal{B}(\widehat{X}),$$

$$c) \quad T(x, \widehat{X}) > 0 \qquad \forall x \in \widehat{X}.$$

We first define T and show that T satisfies c). Set $T(x, A) := P^l(x, A^o)$, for all $x \in \widehat{X}$ and all $A \in \mathcal{B}(\widehat{X})$, where l is chosen such that $P^l(x, \widehat{X}^o) > 0$ for all $x \in X$ (A^o denotes the interior of the set A). Such $l < \infty$ exists because for some finite sequence of environmental cycles the process comes arbitrarily close to the H-period cycle described in Proposition 2, and because this cycle is contained in the interior of $\widehat{X}$. It follows that $T(x, \widehat{X}) = P^l(x, \widehat{X}^o) > 0$, for all $x \in \widehat{X}$, which shows c).

Second, we establish that T defines a sub-stochastic transition kernel, that is,

- $T(\cdot, A)$ is a non-negative measurable function on $\widehat{X}$, for all $A \in \mathcal{B}(\widehat{X})$, and
- $T(x, \cdot)$ is a probability measure on $\mathcal{B}(\widehat{X})$, with the exception that $T(x, \widehat{X}) \leq 1$, for every x.

The first property holds true, because $A \in \mathcal{B}(\widehat{X})$ implies $A^o \in \mathcal{B}(\widehat{X})$ and since $P^l(\cdot, A^o)$ is a non-negative measurable function on $\widehat{X}$. As to the second property, $T(x, A) = P^l(x, A^o) \geq 0$, for all $A \in \mathcal{B}(\widehat{X})$, and $T(x, \widehat{X}) = P^l(x, \widehat{X}^o) \leq 1$ follows from $P^l(x, \cdot)$ being a probability measure on $\mathcal{B}(\widehat{X})$. Hence, it remains to be shown that

$$T\left(x, \bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} T(x, A_i) \tag{A.16}$$

for all pairwisely disjoint sets $\{A_i\}_{i \in \mathbb{N}}$ ($A_j \neq A_k$, for all $j, k \in \mathbb{N}$ such that $j \neq k$). To establish (A.16), notice that

$$\begin{aligned}\sum_{i=1}^{\infty} T(x, A_i) &= \sum_{i=1}^{\infty} P^l_{\widehat{X}}(x, A_i^o) = P^l\left(x, \bigcup_{i=1}^{\infty} A_i^o\right) \\ &= P^l\left(x, \left(\bigcup_{i=1}^{\infty} A_i\right)^o\right) = T\left(x, \bigcup_{i=1}^{\infty} A_i\right)\end{aligned}$$

because of $(\cup_{i=1}^{\infty} A_i)^o = \cup_{i=1}^{\infty} A_i^o$. This shows that $T(x, \cdot)$ constitutes a (sub)probability measure, which in turn implies that T defines a substochastic transition kernel.

Third, $T(\cdot, A)$ is a lower semicontinuous function, because $P^l(\cdot, A^o)$ is a lower semicontinuous function. The latter holds true, since A^o is open and $P^l(\cdot, A^o)$ is a (weak) feller chain on $\widehat{X}$. This completes b). Finally, setting $a := f_l$ (where $f_l : \mathbb{N}_0 \to \{0, 1\}$ such that $f_l(t) = 1$ if $t = l$ and $f_l(t) = 0$ otherwise), we obtain $P^l(x, A) = K_a(x, A) \geq T(x, A) = P^l(x, A^o)$ because of $A^o \subseteq A$, which completes a).

Part iv): $P_{\widehat{X}}$ is Harris recurrent.

According to Tuominen and Tweedie (1979, Theorem 4.2.), $P_{\widehat{X}}$ is Harris recurrent if $a)$ there exists $x_0 \in \widehat{X}$ such that $\inf_{x \in \widehat{X}} P_x \{\tau_N < \infty\} > 0$ for every neighborhood N of x_0, and $b)$ $P_{\widehat{X}}$ is a $T-$chain. Since $a)$ follows from Proposition 2, the claim holds true.

Part v): $P_{\widehat{X}}$ is aperiodic, first, because $\cup_{(i,j) \in J} \widehat{X}_{ij}$ forms a partition of $\widehat{X}$, second, because of Lemma 2 and, third, because the environmental chain is aperiodic.

Part vi): $P_{\widehat{X}}$ is $\varphi-$irreducible.

By Proposition 2, we have $\sum_{t \in \mathbb{N}_0} P^t(x, \mathcal{O}) > 0$, for any $x \in \widehat{X}$ and any neighborhood $\mathcal{O}$ of the H-period limit cycle induced by some arbitrary environmental H-period cycle (i.e., in terminology of Meyn and Tweedie, 1996, all states corresponding to any such limit cycle are *reachable*). Because of $P^t_{\widehat{X}}(x, \mathcal{O}) = P^t(x, \mathcal{O})$ for all $x \in \widehat{X}$ and all $\mathcal{O} \in \mathcal{B}(\widehat{X})$, the claim follows from Proposition 6.2.1 in Meyn and Tweedie (1996, p.133).

Part vii): $P_{\widehat{X}}$ is positive (i.e., it admits an invariant probability measure).

Since $\widehat{X}$ is bounded and topologically closed, $\widehat{X}$ is compact. The claim thus follows from Theorem 12.10 in Stokey and Lucas (1993, p. 376). ■

Department of Economics
University of Bergen
Fosswinckels gate 6
N-5007 Bergen, Norway
Phone: +47 55 58 92 00
Telefax: +47 55 58 92 10
http://www.svf.uib.no/econ