

## Cournot competition and endogenous firm size

Francesco Saraceno · Jason Barr

Published online: 18 July 2008  
© Springer-Verlag 2008

**Abstract** The paper studies the dynamics of firm size in a repeated Cournot game with unknown demand function. We model the firm as a type of artificial neural network. Each period it must learn to map environmental signals to both a demand parameter and its rival's output choice. However, this learning game is in the background, as we focus on the endogenous adjustment of network size. We investigate the long-run evolution of firm/network size as a function of profits, rival's size, and the type of adjustment rules used.

**Keywords** Firm size · Adjustment dynamics · Artificial neural networks · Cournot games

**JEL Classification** C63 · D21 · D83 · L13

---

Earlier versions of this paper were presented at the 10th International Conference on Computing in Economics and Finance, July 8–10, 2004, University of Amsterdam; the International Workshop on Agent-Based Models for Economic Policy Design, Bielefeld, June 30–July 2, 2005; A Symposium on Agent-based Computational Methods in Finance, Game Theory, Lille, September 15–16, 2005. We thank the various session participants for their comments. The authors would also like to acknowledge that partial funding for the use of Compustat was obtained from the David K. Whitcomb Center for Research in Financial Services at Rutgers University.

---

F. Saraceno (✉)  
Observatoire Français des Conjonctures Économiques,  
69 Quai d'Orsay, 75007 Paris, France  
e-mail: francesco.saraceno@sciences-po.fr

J. Barr  
Rutgers University,  
Newark, NJ, USA  
e-mail: jmbarr@rutgers.edu

## 1 Introduction

The dynamics of firm size is an important part of industrial organization. Though there exists a relatively large literature on understanding and modeling the statistical properties of firm growth, there has been little work done on investigating the effect of firms' internal organization on growth.<sup>1</sup> In this paper, we study firm behavior within an agent-based framework to investigate the roles that both firm structure and competition play in firm size dynamics.

We model the firm's internal organization as a type of artificial neural network (ANN). ANNs have many features that make them suitable for simple models of the firm as an information processor (Barr and Saraceno 2002). Here, we have two firms-as-ANNs competing at different levels. At the first level, the networks play a Cournot game, and learn to improve their performance over time by comparing a forecasted best response with the correct best response. The second level of competition is based on firm size, i.e., on the amount of resources that the firm employs in the activity of information processing. Size in this paper is measured by the number of nodes (or agents) in the network; firms try to improve long run profitability by changing their size.

The two forms of competition (Cournot and firm size) are both important components of the model, but in this paper we focus mostly on the dynamics of firm size, while the Cournot game is in the background. Barr and Saraceno (2005, 2008) explored in detail models of Cournot competition between two networks. These models focused on the role that information processing plays in Cournot games. We showed that network size, which was left exogenous, is an important factor in determining learning performance in an uncertain environment.

Here we reverse the perspective, and we take as given both the learning process and the dependence of firm profit on size and environmental complexity, and we endogenize firm size; in other words, we look at the endogenous determination of the resources (i.e., the number of nodes) devoted to learning. The computational power needed to discover the optimal network size (defined as the result of a best response dynamics) can be quite expensive. Thus, this paper adopts the point of view that firms act in a boundedly rational way. Our interest in doing so is to explore to what extent simple rules help us understand firm behavior, given that agents are inherently limited in their capabilities. We also seek to explore the degree to which myopic rules mimic the outcomes of rational choice models. Our main scope is to investigate the role of interaction with rivals in determining firm growth.

Even though firms are multi-agent systems, there are still limits to their organizational capabilities and sizes. They must deal with different types of un-

---

<sup>1</sup>Several authors investigate whether Gibrat's Law of Proportionate effect holds for firms (see Evans 1987; Sutton 1997). Others study the stochastic processes yielding skewed distributions of firm sizes (Ijiri and Simon 1977). Finally there are papers within a stochastic dynamic programming/game theoretic approach (Ericson and Pakes 1995; Jovanovic 1982).

certainty that can not all be properly internalized or neutralized. Furthermore, firms are burdened with the task of coordinating and integrating the different flows of information that emerge from the external environment and also from within the firm itself.<sup>2</sup> In addition, in a complex world, the acquisition of useful information is costly, even for corporate behemoths such as General Motors or Boeing.

In this regard our work fits within the relatively new literature on agent-based models of organizations (see Chang and Harrington 2006, for a review). Organizations are modeled as a collection or network of agents that process incoming data. Information processing (IP) networks and organizations arise because in modern economies no one agent can process all the data. The growth of the modern corporation has created the need for workers who are managers and information processors (Chandler 1977; Radner 1993). Typical models are concerned with the relationship between the structure of the network and its corresponding performance (DeCanio and Watkins 1998; Radner 1993).

In our paper the network uses signals from the economic environment to forecast both demand and a rival's output decision. The firm is able to learn over time as it repeatedly gains experience in observing and making decisions about environmental signals. Unlike other information processing models, our duopoly model explicitly includes strategic interaction: one firm's ability to learn the environment affects the other firm's pay-offs. Thus a firm must locate an optimal network size not only to maximize performance from learning the environment but also to respond to its rival's actions.

A second area of literature that relates to our work is evolutionary and firm decision making theory (Cyert and March 1963; Nelson and Winter 1982; Simon 1976). In this literature, the firm is also boundedly rational, but the focus is not on information processing *per se*. Rather, the firm is engaged in a myriad of activities from production to sales and marketing, R&D, business strategy, etc., which contribute to form a set of capabilities difficult for other firms to replicate. Patterns of behavior ("routines") emerge from the combination of learning by doing, imitation, and R&D (Nelson and Winter 1982). In a world where discovering optimal solutions is computationally expensive, firms will seek relatively easy rules of behavior that produce satisfactory responses at relatively low cost (Simon 1956). In this vein, firms in our paper employ simple rules when adjusting size.

We explore two types of boundedly rational rules in regard to firm size dynamics. The first rule (which we call "isolationist") has the firm adjusting its size simply based on its last period profit growth: a firm adds (removes) information processing agents if profits are growing (shrinking). The isolationist rule is analogous to the common myopic supply adjustment model, where agents adjust their supply based on last period's price (Nerlove 1958).

---

<sup>2</sup>Cyert and March (1963) and Nelson and Winter (1982) discuss how competing agendas of agents within the firm constrain the decisions that can be made.

The second rule (“the imitationist”) has the firm adjusting size by comparison with its rival’s profits and size. If the rival has larger profits, the firm will try to match its size, i.e., increase (decrease) the number of nodes if the rival is larger (smaller). Firm imitation is a quite common phenomena. For example, in the last twenty-five years, there are has been a tremendous surge in the use of management consultants. This allows firms to bring in agents who have knowledge about other firm’s activities and “best practices” (Ernst and Kieser 1999). Furthermore, Mirvis (1997) shows how firms look to their rivals for information about their human resource practices. In his study, 89% of firms surveyed adopted human resource policies based on information collected from other firms: 39% of firms adopted new policies immediately following the “innovation leaders,” 39% of firms adopted new practices after a consensus developed in their respective industries, and 11% adopted new policies and practices only after they proved effective in other companies. In addition, the widespread imitation of technology and business strategies has been well documented in the literature (e.g., Mansfield 1961; Nelson and Winter 1982; Rivkin 2000; Utterback 1994).

Lastly we investigate size dynamics when the firms use a combined rule—using both their own profit and rival’s size and profit to guide their decisions. We explore how firms interact in their respective dynamics, and how the adjustment rate parameters affect long run firm size. We also study the conditions under which the adjustment rules will produce the Cournot–Nash outcome.

To our knowledge, no other paper has developed the concept of long run firm growth in an adaptive or agent-based setting.<sup>3</sup> To anticipate some of our findings: (a) In both the isolationist and the imitationist cases, size dynamics is a nonlinear function of both environmental complexity, initial own and rival’s size, and the adjustment rate parameter. (b) Via a simulation experiment, we give a quantitative measure of the relative effects of initial conditions and parameters on long run size dynamics. Our main conclusion in this case is that interaction between these variables plays a crucial role. (c) We show that the adaptive rules rarely converge to the Nash size. The sensitivity to the firm’s own profits seems to play an important role in moving the firm away from the best response behavior. (d) Lastly, we build a data set of duopolist firms in the U.S. and we use it to provide empirical support for our model.

The next section explains the choice of neural networks as models of the firm. Then, in Section 3 we set up the Cournot model. Section 4 discusses the relationship between profits and firm size, and characterizes the benchmark case of network size equilibria. Sections 5 to 7 discuss the heart of the paper—the firm size adjustment algorithms and the results of the simulations. Section 8 provides an empirical test of our model. Section 9 concludes. An appendix provides additional information about the models.

---

<sup>3</sup>There are other agent-based models of firm growth, such as Axtell (2002), but they do not explore the adaptive dynamics of agent networks.

## 2 Neural networks as models of the firm

In previous work (Barr and Saraceno 2002, 2005) we argued that information processing is an important feature of modern corporations, and that efficiency in performing this task may be crucial for success or failure. We further argued that when focusing on this aspect of firm behavior, computational learning theory may give useful insights and modelling techniques. From this perspective, it is useful to view the firm as a *learning algorithm*, consisting of agents that follow a series of rules and procedures.

Firms learn, and improve their performance by repeating their actions and recognizing patterns (i.e., learning by doing). By processing information, firms learn about the environment and become proficient at recognizing new and related patterns. Furthermore, they face a trade-off linked to the complexity of their organization. Small firms are likely to attain a rather imprecise understanding of the environment they face; but on the other hand they can react quickly and are able to design decent strategies with small amounts of information. Larger and more complex firms produce more sophisticated analyses, but they need time and experience to implement their strategies. Thus, the “optimal” (e.g., profit maximizing) firm structure must be determined in relation to the complexity of the environment, and it is likely to change with it.

Management scholars (such as Chandler 1977; Cyert and March 1963; Lawrence and Lorsch 1986) have described these features of firm behavior in a number of studies. Lawrence and Lorsch (1986), for example, discuss how different types of firms organize themselves differently, based on the nature of the environment and the technology. For example, frozen food manufacturers will have different organizational structures than plastics manufacturers. Even within industries, firm performance can vary. Lawrence and Lorsch found that though plastics manufacturers had similar organizational structures, firm “fitness” was affected by how each firm had adapted to its environment. Rivkin and Siggelkow (2003), demonstrate with an  $NK$  landscape model how firm performance is determined by the relationship of the firm’s internal organization with its environment.

In our previous work we focused on how the number of information processing agents can affect learning performance; in particular, via simulations we showed that the trade-off between speed and accuracy generates a hump-shaped profit curve with firm size (Barr and Saraceno 2002). We also found that optimal firm size is positively related to environmental complexity. These results reappeared when we applied the model to Cournot competition (Barr and Saraceno 2005).

Barr and Saraceno (2002) highlights several parallels between firm learning and machine learning in computational learning theory. Learning machines can generalize from experience to unseen problems, and recognize patterns. Furthermore, they generally face a trade-off between the rapidity of information processing and the accuracy, like the one we described about firms. Finally, a particular type of learning machine, the artificial neural network

(ANN), has an additional interesting feature—parallel and decentralized processing. ANNs are composed of multiple units processing relatively simple tasks in parallel. The combined result of this multiplicity is the ability to process very complex tasks. In the same way firms are often composed of different units working autonomously on very specific tasks, and are coordinated by management that merges the results of these simple operations, creating a total value larger than the sum of the single tasks involved.

### 3 The Cournot game

Here we discuss the Cournot game that the firms play, taken from Barr and Saraceno (2005, 2008). Two firms face a downward sloping, linear demand,  $p_t = \gamma_t - (q_{1t} + q_{2t})$ , where  $\gamma_t$  changes over time and is *ex ante* unknown. We assume that production costs are zero and the slope is fixed and normalized to one.  $\gamma_t$  is a function of a set of commonly observable environmental variables  $\mathbf{x} \in \{0, 1\}^n$ :

$$\gamma_t = \gamma(\mathbf{x}_t) = \frac{1}{(2^n - 1)} \sum_{k=1}^n x_{kt} 2^{n-k}. \quad (1)$$

$\gamma(\mathbf{x}_t)$  can be interpreted as a weighted sum of the presence or absence of environmental features. A simple measure of environmental complexity is  $n$ , the number of bits in the vector  $\mathbf{x}$ .<sup>4</sup> Assuming that  $\gamma_t$  were known, we would have the following best response function:

$$q_i^{br} = \frac{1}{2} [\gamma - q_{-i}], \quad i = 1, 2,$$

(for notational convenience time subscripts are used only when needed). The symmetric Nash equilibrium output and corresponding profit would then be

$$q^{ne} = \frac{\gamma}{3}, \quad \pi^{ne} = \frac{\gamma^2}{9}.$$

Because costs in this case are zero, the quantity produced and profits are the same for both firms. Note that in standard models of Cournot competition, equilibrium output is sometimes interpreted as firm size, which is generally a function of production costs. To avoid confusion it is worth repeating that in our setting firm size is instead related to the learning and information processing activity. Firms of different (network) sizes can thus produce the same quantity, if their estimate of the uncertain environment coincides.

<sup>4</sup>Notice that  $n$  encompasses two measures of environmental complexity. First, more bits in the vector means more information to process, *each period*. Secondly,  $n$  also determines (inversely) the probability that a particular vector will be selected, i.e., the total amount of information to process, *over time*.

In particular, uncertainty in this model is related to Eq. 1. Firms only know that  $\gamma_t$  depends on  $\mathbf{x}_t$ , but they have to learn the mapping between the two. Each period, they view an environmental vector  $\mathbf{x}_t$ , which is determined by random draws from a uniform distribution (i.e.,  $Pr(\mathbf{x}_t = \mathbf{x}) = 1/2^n$ ), and they use this information to estimate the value of  $\gamma(\mathbf{x}_t)$ .

As mentioned above, we model firms as artificial neural networks (ANNs), which are nonlinear approximators that can map virtually any function. Their flexibility makes them powerful tools for pattern recognition, classification, and forecasting. In particular, we use the most popular network architecture, the backward propagation network (BPN). Here we give a brief description of how they work, with more analytic details given in Appendix A.

The network consists of an “input layer” (in our case the environmental vector  $\mathbf{x}_t$ ), and a collection of nodes organized in “hidden layers,” representing the managers that process information in the organization. The last layer is the output/decision layer. Inputs and nodes are connected by weights, that store the knowledge of the network. The inputs are passed through the neural network to determine an output. Each agent in a given layer takes a weighted sum of the signals from agents in the layer below, and then applies a squashing function  $g(\cdot)$  (e.g., the sigmoid), to allow for nonlinear transformations; the agent then passes the processed information to the layer above. We use a single hidden layer network so that the information processing activity of the firm can be represented by:

$$\hat{\mathbf{y}} = g[g(\mathbf{x} \cdot \mathbf{w}^h) \mathbf{w}^o],$$

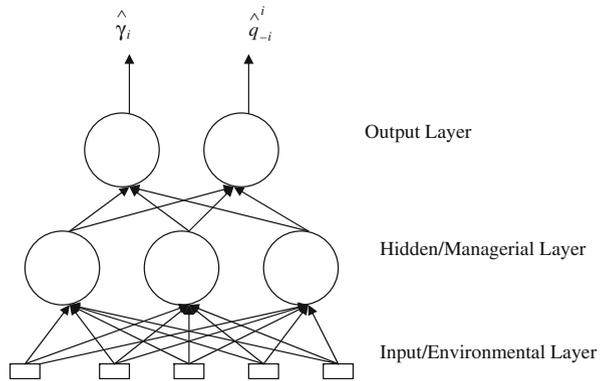
where  $\hat{\mathbf{y}} \equiv \{\hat{\gamma}^i, \hat{q}_{-i}^i\}$ ,  $\mathbf{w}^h$  and  $\mathbf{w}^o$  are the weight matrices. Notice that the squashing function  $g(\cdot)$  is applied twice (i.e., first from the inputs to the hidden layer, and then from the hidden layer to the output layer). Figure 1 presents the graph of an ANN.

The network learns via successive adjustments of the weights, with the objective of minimizing a (squared) error. Supervised learning takes place in the sense that at each iteration the network output is compared with a known correct answer, and weights are adjusted backward in the direction that reduces the error (the gradient descent method). The learning process is stopped once it attains a threshold level for the error, or a fixed number of iterations has elapsed. Thus, in each period:

1. Firms observe a randomly drawn environmental state vector  $\mathbf{x}_t$ .
2. Based on that, each firm estimates a value of the intercept parameter,  $\hat{\gamma}^i(\mathbf{x}_t)$  and its rival's choice of output,  $\hat{q}_{-i}^i(\mathbf{x}_t)$ , where  $\hat{q}_{-i}^i$  is firm  $i$ 's guess of firm  $-i$ 's output.<sup>5</sup> Firms then make an output choice based on the best response function:  $q_i^{br} = \frac{1}{2} [\hat{\gamma}^i - \hat{q}_{-i}^i]$ .

<sup>5</sup>The firm uses environmental information to guess the rival's output choice. Since we assume that both firms are learning to play a best response, this is equivalent to “learning the learning process” of the rival.

**Fig. 1** Graph of an artificial neural network



3. Firms then observe the true value of  $\gamma$  and  $q_{-i}$ , and compute their forecast errors:

$$\begin{aligned} \varepsilon_{\gamma i} &= (\hat{\gamma}^i - \gamma)^2 \\ \varepsilon_{q_i} &= (\hat{q}_{-i}^i - q_{-i})^2 \end{aligned} \tag{2}$$

4. Based on these errors, firms update the weight values in its network. The details of the learning algorithm are given in Appendix A.

These steps are repeated for  $T = 250$  iterations. At the end, we compute the average profit for the two firms as

$$\pi_i = \frac{1}{T} \sum_{t=1}^T \pi_{it} = \frac{1}{T} \sum_{t=1}^T q_{it} (\gamma_t - (q_{1t}^{br} + q_{2t}^{br})). \tag{3}$$

Average profit is a good measure of firm performance since it captures how well the firm does in repeated plays of the Cournot game. As discussed above, large networks are slow to learn, but can achieve a better accuracy over time; smaller networks are much quicker in their learning, but in the long run they are not as accurate. Average profit allows us to determine which types of firms best balance the trade-off between speed and accuracy.

Note that our setting is different from Cournot models such as Novshek and Sonnenschein (1982) and Kirby (1988), where Cournot duopolists must determine how much information about a stochastic demand should be shared, versus how much should be kept private. Here, both firms view a common set of environmental signals, which they use to help determine their quantity decisions. As such, there is no concern about equilibrium levels of information sharing. Rather, the only equilibrium notion embedded here is that each firm seeks to play a best response given its forecasts of the intercept and of the rival’s output; and therefore, both firms are moving toward the Nash output equilibrium.

Profits can be shown to depend on the errors of both firms, and on the model fundamentals. If the two firms exactly forecasted the correct intercept and the

rival’s outputs, we would obtain the Cournot quantity and profit ( $q^{ne} = \gamma/3$ ,  $\pi^{ne} = \gamma^2/9$ ). We can calculate the “loss” or deviation of actual from optimal profits, caused by imprecise forecasts, as

$$L_i = \pi^{ne} - \pi_i = p^{ne}q^{ne} - pq_i, \quad i = 1, 2. \tag{4}$$

With some manipulation we obtain  $L_i = p^{ne}q^{ne} - pq_i + (p^{ne}q_i - p^{ne}q_i) = \frac{\gamma}{3}(\frac{\gamma}{3} - q_i) + q_i(\frac{\gamma}{3} - p)$ , where  $q_i$  is the best response given the forecasted values of  $\gamma$  and  $q_{-i}$ :

$$q_i = q_i^{br} = \frac{1}{2}(\hat{\gamma}^i - \hat{q}_{-i}). \tag{5}$$

From the definitions of errors (Eq. 2) we obtain

$$\hat{\gamma}^i = \sqrt{\varepsilon_{\gamma i}} + \gamma, \quad \hat{q}_{-i} = \sqrt{\varepsilon_{q_{-i}}} + q_{-i}. \tag{6}$$

Putting together Eqs. 5 and 6, we have:

$$q_i = \frac{1}{2}(\psi_i + \gamma - q_{-i}), \tag{7}$$

where  $\psi_i \equiv (\sqrt{\varepsilon_{\gamma i}} - \sqrt{\varepsilon_{q_i}})$ . We plug  $q_{-i} = \frac{1}{2}(\psi_{-i} + \gamma - q_i)$  into Eq. 7, rearrange, and solve for  $q_i$

$$q_i = \frac{1}{3}(\gamma + 2\psi_i - \psi_{-i}). \tag{8}$$

Notice that if all errors are zero, then  $q_i = \gamma/3 = q^{ne}$ . We can plug  $q_i$  and  $q_{-i}$  into the price function to obtain:

$$p = \frac{1}{3}(\gamma + \psi_1 + \psi_2). \tag{9}$$

Plugging the price and quantities (Eqs. 8 and 9, respectively) into the loss function (4) we obtain  $L_i = \frac{1}{9}(\psi_{-i}^2 - \psi_i\psi_{-i} - 2\psi_i^2 - 3\gamma\psi_i)$ . Therefore, profit can be written as the deviation from optimal profit:

$$\pi_i = \pi^{ne} - L_i = p^{ne}q^{ne} - \frac{1}{9}(\psi_{-i}^2 - \psi_i\psi_{-i} - 2\psi_i^2 - 3\gamma\psi_i).$$

Notice that loss tends to 0 as the errors tend to 0, but also that it needs not to be positive. The forecast errors may result in a profit larger than the Cournot-Nash level for one (but not both) firms. Think for instance of the case in which the combined production is equal to the Cournot level, but firm  $i$  produces more than the other. Firm  $i$  will then have a profit larger than the equilibrium one (as  $p = p^{ne}$ , and  $q_i > q^{ne}$ ).

### 4 Firm learning and profits

Barr and Saraceno (2005) show that, in general, firms-as-neural networks are capable of learning how to map environmental factors to demand and to rival’s output; over time this allows them to converge to the Nash equilibrium. We

further show that the main determinants of firm profitability are firm sizes (i.e., the number of processing units,  $m_1$  and  $m_2$ ), and environmental complexity, (i.e., the number of inputs to the network,  $n$ ). These facts may be succinctly captured by a polynomial in the three variables:

$$\pi_i = f(m_1, m_2, n) \quad i = 1, 2. \tag{10}$$

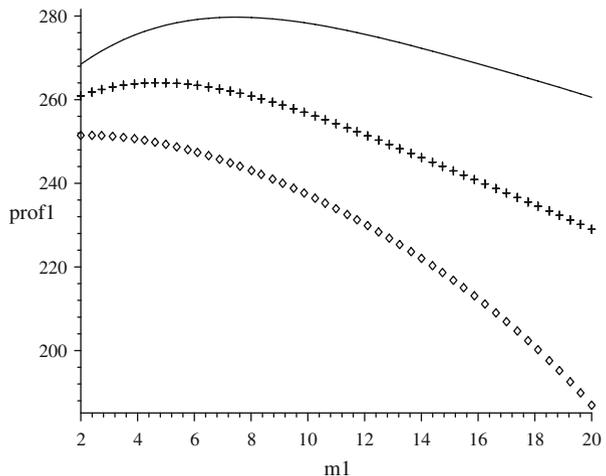
To obtain a specific numerical form for Eq. 10, we simulated the Cournot learning process with different randomly drawn firm sizes ( $m_1, m_2 \in [2, 20]$ ) and complexity values ( $n \in [5, 45]$ ), recording each time the average profit of the two firms. Doing this, we obtained an artificial data set, relating average profit to firm sizes and complexity. With this data set we ran a regression whose results are reported in the appendix (Table 5). The regression coefficients are then used to produce a profit equation:

$$\begin{aligned} \pi_1 = & 271 + 5.93m_1 - 0.38m_1^2 + 0.007m_1^3 + 0.49m_2 \\ & - 0.3m_1m_2 - 2.2n + 0.0033n^2 + 0.007m_1m_2n - 0.016m_2n. \end{aligned} \tag{11}$$

Notice that the setup is symmetric, so that either firm could be used. Figure 2 shows the profit equation (11) with respect to  $m_1$ , holding the other variables fixed. It shows that profits are hump shaped with respect to own size. Three curves are reported, corresponding to small ( $m_2 = 2$ ), medium ( $m_2 = 10$ ) and large ( $m_2 = 20$ ) opponent’s size (complexity is fixed at  $n = 10$ ).

It is interesting to notice that the trade-off between speed and accuracy appears regardless of the opponent’s size, giving a humped-shape relationship between profit and own size. It is in fact an inherent feature of the learning process studied in the field of computational learning theory (Niyogi 1998). It is also worth mentioning that profit is decreasing with the opponent’s size (the curve shifts down as  $m_2$  increases).

**Fig. 2** Firm 1 profit function vs. own size  $m_1$  ( $m_2 = 2$ , solid line;  $m_2 = 10$ , crosses;  $m_2 = 20$ , diamonds).  $n = 10$



#### 4.1 The best response function and network size equilibrium

In this section, as a benchmark, we discuss the firm's best response function and equilibria with respect to firm size. Broadly speaking, we can think of the Cournot game as happening on a short term basis, while the adjustment dynamics occurs over longer periods of time. We can derive the best response function in size by setting the derivative of profit with respect to size equal to zero, i.e.,

$$\frac{\partial \pi_i}{\partial m_i} = \frac{\partial f(m_i, m_{-i}, n)}{\partial m_i} = 0 \quad i = 1, 2.$$

Given the functional form of Eq. 11, this yields the following solution for firm  $i$ :<sup>6</sup>

$$m_i^{br}(m_{-i}, n) = 16.9 \pm 2.26\sqrt{2.6m_{-i} - 0.058nm_{-i} + 3.9}.$$

The “best response” function is polynomial in  $m_i$ . Depending on the particular values of  $m_{-i}$  and  $n$ , we can have up to two real solutions, or two complex solutions. For values of  $m_{-i}$  and  $n$  in the admissible range ( $m_i, m_{-i} \in [2, 20]$ ,  $n \in [5, 45]$ ), the best response is real and decreasing in  $m_{-i}$ .<sup>7</sup>

We define the *network size equilibrium* (NSE) as the intersection of the two firms' best responses. Figure 3 shows the best response mappings for Eq. 11, and the corresponding Nash equilibria in size, along the 45 degree line. Notice that these equilibria are stable, and that increasing complexity shifts upwards the best response functions, and consequently the Nash equilibrium size.

We can express optimal firm size,  $m_i^*$ , and the corresponding profit  $\pi_i^*$ , as functions of environmental complexity. Taking the best responses, and imposing  $m_i^* = m^*$  by symmetry, we obtain the following increasing relationship:

$$m^* = 23.5 - 0.15n - 4.5\sqrt{14.1 - 0.34n + 0.001n^2} \quad (12)$$

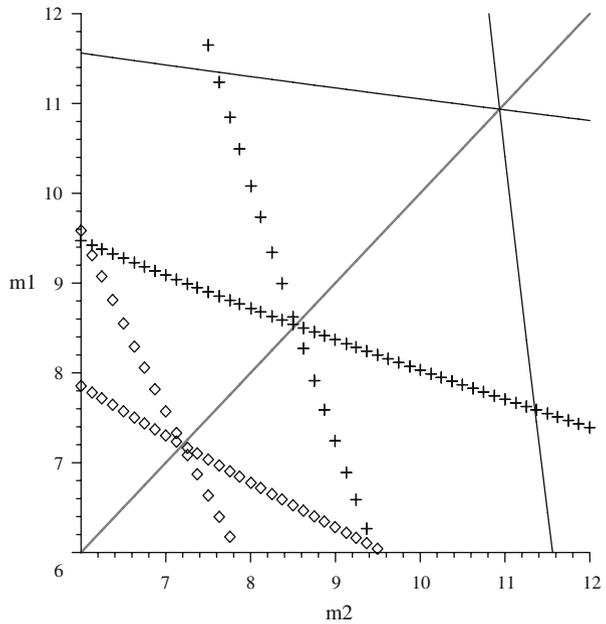
Finally, substituting the optimal value given by Eq. 12 into the profit equation 11, we obtain a decreasing relationship.  $m^*(n)$  and  $\pi^*(n)$  are plotted in Fig. 4.<sup>8</sup> The shape of these curves is rather intuitive. As  $n$  increases, *ceteris paribus*, the learning process becomes less accurate; firm errors increase and therefore profits, which are a negative function of the errors, decrease (see Barr and Saraceno 2005). The optimal size increases as well, as the optimal balance between speed and accuracy, in more complex environments, requires more computational power. To sum up, we have shown that best response dynamics yield a unique and stable equilibrium in firm size. This is increasing in complexity, while profit is decreasing.

<sup>6</sup>For simplicity we ignore the integer issue and assume that firm size can take on any real value in the  $[2, 20]$  interval.

<sup>7</sup>For each value of  $m_{-i}$  and  $n$  we have two positive and real roots for  $m_i$ . However, we discard the larger one since it is outside the admissible range.

<sup>8</sup>As was the case before, we actually have two solutions for firm size. Nevertheless, one root gives values that are outside the relevant range for  $m^*$ , and can be discarded.

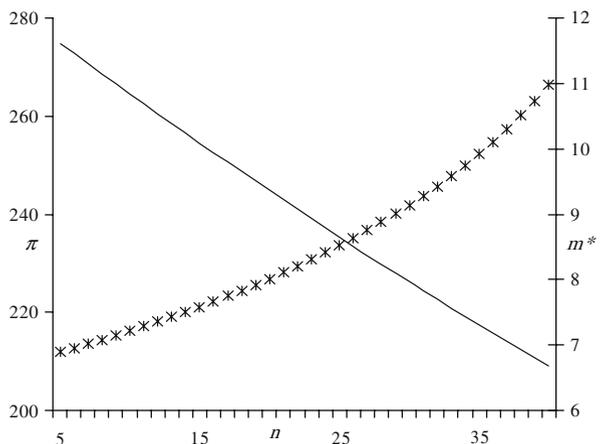
**Fig. 3** Best response functions ( $n = 10$ , diamonds;  $n = 25$ , crosses;  $n = 40$ , solid lines). The NSE are given by the intersections



### 5 Adaptive adjustment dynamics

In standard economic theory, firms are generally assumed to understand many details about how the world functions. They know their cost and profit functions, and the effect of a rival's decisions or environmental variables on profits. The best response function in Section 4.1 is quite complex, and firms are assumed to know the expected maximal profit obtainable for the entire range of a rival's choice of network size.

**Fig. 4** Equilibrium profit (solid line, left axis) and equilibrium size (crosses, right axis) as functions of environmental complexity



But, in a complex world, the cost to discover such knowledge is relatively large. As a result, firms engage in routine behavior, and use rules of thumb to guide their action (Nelson and Winter 1982). In addition, in a world in which production and market conditions constantly change, past information may quickly become irrelevant: even if a firm has perfect knowledge of its best response function at a certain point in time, that knowledge may soon become outdated. That means that even when the firm has the computational capabilities necessary to compute the best response, it may not find it efficient to actually do so. For these reasons, in this section, we explore simple dynamics for firm size. We assume that firms observe their and their rival’s profits and sizes, and we explore adjustment dynamics using the following rule:

$$m_{i,t} = m_{i,t-1} + \beta (\pi_{i,t-1} - \pi_{i,t-2}) + \alpha I_i [(m_{-i,t-1} - m_{i,t-1})(\pi_{-i,t-1} - \pi_{i,t-1})]. \tag{13}$$

$\beta \geq 0$  represents the sensitivity of firm size to its own profit growth. If a firm has positive profit growth it will increase its size by  $\beta (\pi_{i,t-1} - \pi_{i,t-2})$  units. The parameter  $\alpha \geq 0$  captures the “imitation” factor behind size adjustment;  $I_i$  is an indicator function taking the value of 1 if the opponent’s profit is larger, and a value of 0 otherwise:

$$I_i = \begin{cases} 1 & \text{if } (\pi_{-i,t-1} - \pi_{i,t-1}) > 0 \\ 0 & \text{if } (\pi_{-i,t-1} - \pi_{i,t-1}) \leq 0. \end{cases}$$

Thus, the firm will adjust towards the opponent’s size, whenever it observes a better performance of the latter ( $I_i = 1$ ). To sum up, our adjustment rule only uses basic routines: first, the firm expands if it sees its profit increasing; second it adjusts towards the opponent’s size whenever it sees that the latter is doing better. These routines are relatively simple, and require very little observation and computation on the part of the firm. In the next section we investigate the firm dynamics yielded by the simple adjustment rules.

## 6 Results

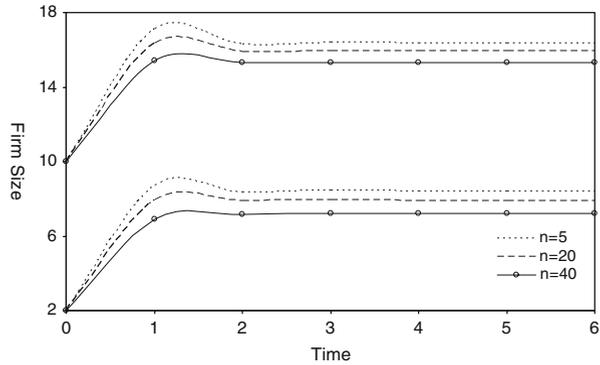
### 6.1 Scenario 1: the isolationist firm

If  $\alpha = 0$ , each firm will only look at its own past performance when deciding whether to add or to remove nodes:

$$m_{i,t} = m_{i,t-1} + \beta [\pi_{i,t-1}(m_{i,t-1}, m_{-i,t-1}, n) - \pi_{i,t-2}(m_{i,t-2}, m_{-i,t-2}, n)]. \tag{14}$$

Of course this does not mean that the firm’s own dynamics is independent of the other, since  $\pi_{i,t-1}$  and  $\pi_{i,t-2}$  depend on both firms’ sizes. Figures 5 and 6 show the adjustment dynamics of two isolationist firms. In the first case we show the evolution for different complexity levels (keeping the adjustment rate parameter fixed at  $\beta = 0.025$ ). In the second case, we keep complexity constant ( $n = 20$ ) and we study dynamics corresponding to different values of  $\beta$ .

**Fig. 5** Firm dynamics when firms start at different sizes ( $m_1(0) = 2; m_2(0) = 10$ ), for three different levels of complexity.  $\beta = 0.025$



The dynamics are in both cases relatively simple. First, the long-run level, in general, is reached fairly quickly. Second, this level depends on initial own size. For example, in Fig. 5, a firm starting at 10 nodes will converge at a value around 16-17 nodes, while a firm starting at 2 will converge to 6-7. Complexity seems to have a role, as it yields lower long run sizes. As for the effects of the adjustment parameter  $\beta$ , we see that too large values may trigger oscillatory dynamics: too large values of size yields profit fluctuations, that in turn feed back with a lag in size changes. The effect of the opponent's initial size is impossible to assess from the figures. Since these single time series are just examples, in Section 6.3 below we generate a large sample of observations from time series simulations, and show regression results for the dynamics of Eq. 13.<sup>9</sup>

## 6.2 Scenario 2: the imitationist firm

The imitationist firm does not care about its own situation ( $\beta = 0$ ), but compares its performance to the opponent:

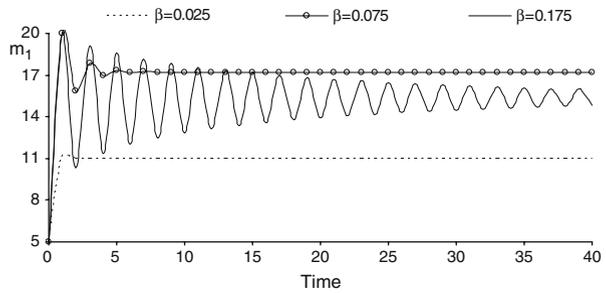
$$m_{i,t} = m_{i,t-1} + \alpha I_i [(m_{-i,t-1} - m_{i,t-1})(\pi_{-i,t-1} - \pi_{i,t-1})]. \quad (15)$$

The firm will not change size if it has a larger profit, and it will adjust towards the opponent if it has a smaller profit.<sup>10</sup> Thus, at each period only one firm moves; further, at equilibrium the two profits must be equal, which happens when firm sizes are equal, but not necessarily only in this situation. For low complexity ( $n = 5$ ), suppose one firm begins small and the other large ( $m_1 = 4$  and  $m_2 = 15$ ). The resulting dynamics depend on  $\alpha$ , as shown in Fig. 7. For low values of  $\alpha$ , in fact, the drive to imitation is not important enough, and firms do

<sup>9</sup>Note that when we choose “extreme” parameters we may get dynamics with firm sizes greater than 20, outside the admissible sample range. For this reason we cap firm size at 20. This can be justified by the fact that at any given time, technology, competition, and human capability will put limits on how big even the largest firm can get. A shorter, companion paper, Barr and Saraceno (2006) shows simulation for some of the cases without caps. The qualitative results do not change.

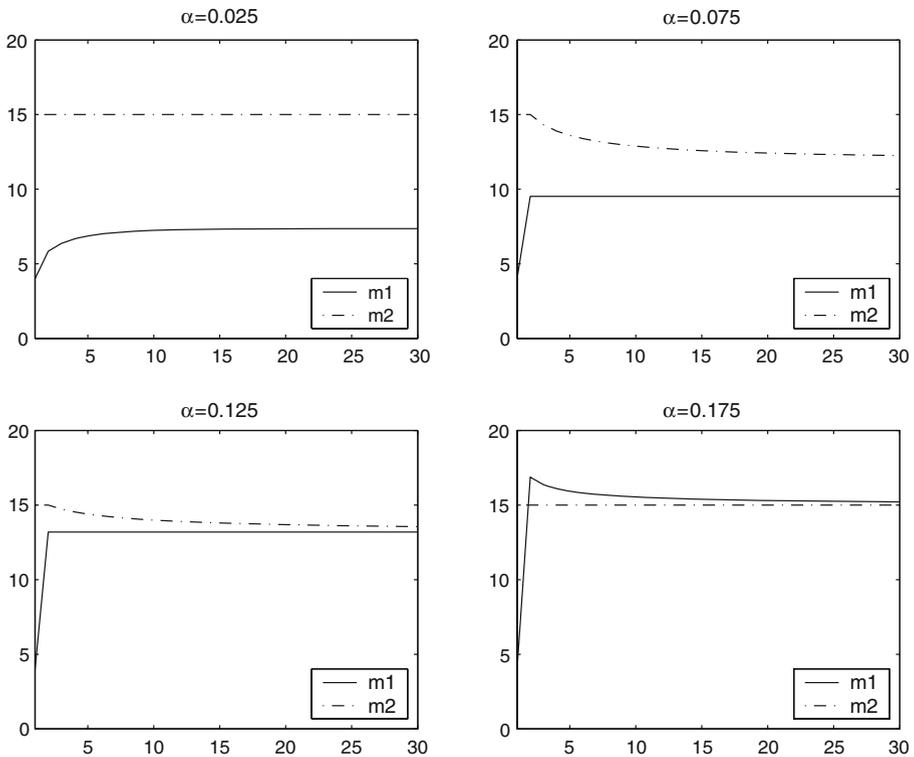
<sup>10</sup>In the special case that both firms have the same initial size there will, of course, be no adjustment.

**Fig. 6**  $m_1(0) = m_2(0) = 5$ ,  $n = 20$ , three different  $\beta$ 's



not converge. For intermediate values (around  $\alpha = 0.125$ ) convergence takes place to an intermediate size. When  $\alpha$  is too large, on the other hand, the initial movement (in this case for firm 1) is excessive, and may overshoot. By increasing complexity we see faster adjustment and, for large enough values of  $\alpha$ , the system explodes (figures are available upon request).

Both in the imitationist and the isolationist cases, the dynamics show a very strong dependence on initial conditions. This feature of the time series calls for a systematic analysis of the parameter space, which we perform in the next section through a simulation experiment.



**Fig. 7** Firm dynamics for different values of  $\alpha$ .  $n = 5$

**Table 1** Dependent variable:  $m_1(100) - m_1(0)$ 

Variable	Coef.	Variable	Coef.
<i>Con</i>	-2.854 (0.35)	<i>dist</i>	-0.468 (0.144)
$[m_1(0)]^2$	0.228 (0.025)	<i>dist · dum</i>	-0.29 (0.047)
$[m_1(0)]^3$	-0.015 (0.001)	<i>n</i>	0.034 (0.004)
$[m_1(0)]^4$	0.0003 (3.98E-05)	$\alpha$	52.2 (2.81)
$m_2(0)$	0.597 (0.149)	$\alpha \cdot m_1(0)$	-3.84 (0.268)
$[m_2(0)]^3$	-0.0007 (5.85E-05)	$\beta$	243.2 (7.659)
$m_1(0) m_2(0)$	-0.189 (0.009)	$\beta^2$	-399.6 (77.8)
$[m_1(0) \cdot m_2(0)]^2$	0.0006 (2.94E-05)	$\beta \cdot m_1(0)$	-4.755 (0.41)
$[m_1(0) \cdot m_2(0)]^3$	-6.9E-07 (4.62E-08)	$\beta \cdot n$	-0.68 (0.11)
<i>nobs</i>	3000		
$R^2$	0.951		

Regression results for adjustment dynamics. Robust standard errors given. All variables stat. sig. at 99% or greater confidence level

### 6.3 Combined dynamics: a simulation experiment

Here, we investigate via a simulation experiment the effects of initial firm size, adjustment parameters and complexity, on long run firm size. We combine the two types of dynamics, given by Eq. 13. To do this, we generate 3,000 random combinations of  $m_i(0) \in [2, 20]$ ,  $\alpha, \beta \in [0.025, 0.075]$ ,  $n \in \{5, 10, \dots, 45\}$  and use the parameters to run the simulation, for 30 iterations. To isolate the effect of the different variables on long run size (which ranges from 2.6 to 20.0 nodes), we use firm growth ( $m_1(100) - m_1(0)$ ) as the dependent variable, and regress it on the parameters.

The regression (Table 1) is non-linear and includes several interaction terms.<sup>11</sup> We also included a dummy variable,  $dum = 1$  if  $m_1(0) > m_2(0)$ , and a distance variable  $dist = m_1(0) - m_2(0)$ . The regression coefficients allow us to isolate the various effects: (a) Initial sizes interact in determining long run growth: a larger opponent ( $m_2(0)$ ) yields negative growth over the period, unless  $m_1(0)$  is also large. In fact, own size has a positive effect (large firms tend to grow larger), that is stronger than the effect of the opponent. There is a general tendency to convergence (negative coefficient for *dist*), even more marked when firm 1 starts larger (the interaction with *dum* is also negative). (b) The parameters  $\alpha$  and  $\beta$  have a positive direct effect, mitigated by the interaction with initial size, and for  $\beta$ , with complexity. In the case of  $\alpha$  the interaction effect is strong enough to reverse the effect for large initial own

<sup>11</sup>Appendix B describes the method for selecting the functional form for the regression.

size. (c) Increasing complexity has a positive direct effect on firm growth. The interaction with  $\beta$  has negative effects, but it is more than compensated by the direct effect of  $\beta$ .

To conclude, the joint action of the simple rules tends to yield reasonable dynamics, and mitigate some of the apparent contradictions that we highlighted when analyzing the rules separately. Increasing complexity tends to increase firm size, whereas a larger opponent size tends to decrease it. This is consistent with the best response in size analyzed in Section 4.1.

### 7 Convergence to Nash size equilibrium

As we saw in the previous section, the long run size of the firm is determined by several variables. This section investigates the conditions under which the simple adaptive size dynamics converge to the Nash equilibrium. To do this we generated a data set of 60,000 observations by taking random draws of the relevant parameters  $m_i(0) \in [2, 20]$ ,  $\alpha, \beta \in [0.025, 0.075]$ ,  $n \in \{5, 10, \dots, 45\}$  and then let the two firms compete for 30 periods. We created a new variable *HitNash* that for each run took the value of 1 if both firms came within  $\pm 0.5$  nodes of the Nash equilibrium size, and 0 otherwise.

Using *HitNash*, as our dependent variable, we ran a probit regression, which measures how the parameters affect the probability that the system will reach the equilibrium. The results are given in Table 2, where we report the results of the marginal changes in probability with a change in the independent variable. For this exercise, the right hand side was generally non-linear in the parameters. In addition we included the dummy variable *dinit*, equal to 1 if the initial sizes are the same ( $m_1(0) = m_2(0)$ ), and 0 otherwise.

**Table 2** Probit regression. Dep. var. is 1 if long run firm size is equal to Nash equilibrium, 0 otherwise. Robust standard errors in parentheses. All coefficients are stat. sig. at 99% or greater confidence level, except  $\alpha$  and  $\alpha \cdot \beta$ , which are stat. sig. at 98.7% and 97.6%, respectively

Variable	df/dx	Variable	df/dx
$n$	0.00031 (0.00005)	$m_2(0)$	-0.00012 (0.00002)
$n^2$	-0.00014 (2.4E-06)	$\beta \cdot m_1(0)$	0.00071 (0.00014)
$n^3$	1.98E-07 (3.6.E-08)	$\beta \cdot m_2(0)$	0.00074 (0.00015)
$\beta$	-0.088 (0.014)	<i>dinit</i>	0.00040 (0.0002)
$\beta^2$	0.991 (0.152)	$\alpha$	0.0019 (0.0009)
$\beta^3$	-2.79 (0.42)	$\alpha \cdot \beta$	-0.0137 (0.006)
$m_1(0)$	-0.00012 (0.0002)	$\beta \cdot n$	-0.00065 (0.0001)
Pseudo R2	0.369	nobs.	60000
obs. Prob.	0.0171	pred.prob (at $\bar{x}$ )	0.00028

The probability of convergence to a Nash equilibrium is quite low—1.7% for our sample. This shows that the simple adjustment rules are generally not equivalent to a best response dynamics. Looking at the probit coefficients of Table 2, we can draw a few conclusions: (a) Larger  $\alpha$  is associated with a higher probability of reaching Nash, despite the negative interaction with  $\beta$ ; the effect of  $\beta$  is negative in the relevant range. This is due to the fact that the  $\alpha$  and  $\beta$  respectively reflect the firm’s “outward” and “inward” orientation, and thus may be seen as indicators of closeness to a best response dynamics. The larger is  $\alpha$ , the greater are the chances to mimic the best response dynamics, whereas the contrary holds for  $\beta$ . (b) The larger the initial firm size, the lower the probability of hitting the Nash equilibrium (despite positive interaction with  $\beta$ ). This is explained by the tendency of large firms to grow larger, which we observed above. (c) The tendency of firm size to grow explains also why we find a positive coefficient for the complexity coefficients. In fact, higher complexity yields larger Nash size.

## 8 Empirical validation

This section uses the Standard and Poor’s Compustat data set to provide a rough empirical assessment of our model. The database contains standardized accounting data for thousands of publicly traded US firms. In this section our measure of firm size is simply the number of employees in the firm, which is an approximate proxy for the number of information processing agents.

We selected our sample of firms from the North American Annual Industrial database. We first downloaded annual data for all active firms (i.e., still in existence as of December 2005) from 1990 to 2005 that had (1) at least one million dollars in sales, (2) a positive value for property, plant and equipment, (3) at least one employee, (4) positive research and development expenditures and (5) data on operating income before depreciation (OIBD), which is our measure of firm profits. OIBD is a measure of the net cash flows from business operations. It is less volatile, in general, than net income, which may include, for example, revenues and costs from a business’ non-core operations.

**Table 3** Descriptive statistics for “duopoly” firms, 1990–2005, 810 observations

Variable	Mean	Std.	Min.	Max.
Sales <sub><i>t</i></sub> (\$Mil.)	2,137	5,110.3	1	62,726
OIBD <sub><i>t</i></sub> (\$Mil.)	257.20	778.9	−981	15,467
Prop., Plant & Equip.(PPE <sub><i>t</i></sub> ) (\$Mil.)	757.79	2,260.8	0	22,840
Emp. Change ( $\Delta$ Emps <sub><i>t</i></sub> ) (000)	0.091	1.79	−9.1	12.9
$\Delta$ OIBD <sub><i>t</i></sub>	17.2	224.2	−3041	3,326
$\Delta$ Sales <sub><i>t</i></sub>	128.68	648.1	−5,189.4	6,242
R&D <sub><i>t</i></sub> /Sales <sub><i>t-1</i></sub>	0.186	2.1	0	50
Dummy riv. bigger/more prof. (RB <sub><i>t</i></sub> )	0.384			
Dummy riv. smaller/more prof. (RS <sub><i>t</i></sub> )	0.110			

The sample contained industries with different competitive structures. In order to obtain a data set relevant for the empirical validation of our model, we restricted the sample to firms that were in “duopoly” industries in the following (imperfect) sense: for each industrial category (defined using the 4-digit standard industrial classification [SIC] code) we dropped from the sample all the observation for which there were more (or less) than two businesses in the industry in a particular year. Once we were left with “duopoly” observations, our sample allowed us to gather information about firms and about their “rivals.”

We eliminated extreme outliers by removing firms for which the dependent variable  $\Delta \text{Emp}_t$  (change in the number of employees from one year to the next) was in either the bottom or top one percentile of firms. Table 3 gives the descriptive statistics of the variables. The three equations to test are given by our model (Eqs. 13, 14 and 15):

$$\begin{aligned}\Delta m_{i,t} &= \beta (\pi_{i,t-1} - \pi_{i,t-2}) + \alpha I_i [(m_{-i,t-1} - m_{i,t-1})(\pi_{-i,t-1} - \pi_{i,t-1})] \\ \Delta m_{i,t} &= \beta (\pi_{i,t-1} - \pi_{i,t-2}) \\ \Delta m_{i,t} &= \alpha I_i [(m_{-i,t-1} - m_{i,t-1})(\pi_{-i,t-1} - \pi_{i,t-1})].\end{aligned}$$

The basic econometric model is given by

$$y_{it} = a_i + b y_{it-1} + \mathbf{c}' \mathbf{x}_{it} + \mathbf{d}' \mathbf{z}_{it} + e_t + \mu_{it}. \quad (16)$$

$a_i$  is a firm-specific constant.  $x_{it}$  is a vector of observable control variables, necessary to avoid omitted variable bias;  $z$  is the vector of variables that we are interested in testing, which are given by Eqs. 13, 14 and 15;  $e_t$  are year dummy variables and  $\mu_{it}$  is the random error. The variable  $y_{it-1}$  is included to help control for unobserved variables, whose omission may bias the coefficients we are interested in measuring. A lagged dependent variable is particularly useful when there is inertia from year to year (and this is true here with employment growth, see Nickell 1996).

To eliminate the firm fixed-effect we take differences of the dependent variable, but this introduces correlation between the lag dependent variable and  $\mu_{it}$ . In addition, some of our control variables may be endogenous due their possible joint determination with the dependent variable. To account for these effects, we employ a method-of-moments estimator (Arellano and Bond 1991), which uses lags of the right-hand side variables as instruments. Table 4 presents the results of our regressions, which only lists the coefficient estimates for variables that test our simulation model.<sup>12</sup>

<sup>12</sup>For all our equations we use as controls the natural log of firm sales, the change in firm sales, the lag of the value of firms property, plant and equipment (i.e., the capital stock), and research and development expenditures divided by the lagged sales (R&D intensity). R&D intensity is our proxy for industrial complexity: presumably firms that have high R&D expenditures are in industries marked by rapid technological change and strong technology-based competition. We also include dummy variables for the year and for the month that each firm ends its fiscal year. Regressions were performed in Stata 9.0; full regression results are available upon request.

**Table 4** Dependent variable:  $\Delta\text{Emp}_t$ . Absolute value of robust z-statistics in parentheses, \*significant at greater than 95%; \*\*stat. sig. at greater than 99%. The P-value for the Sargan test for over-identifying restrictions is equal to 1.00 for all three equations

	(1)	(2)	(3)
$\Delta\text{OIBD}_{t-1}$	0.004 (2.95)**		0.004 (2.53)*
$\text{RB}_{t-1}$		1.17 (1.48)	1.22 (1.53)
$\text{RS}_{t-1}$		-1.90 (2.32)*	-1.60 (2.09)*
Obs.	264	264	264
Wald $\chi^2$	5,161	371	6,672

To test the isolationist equation, we regress the change in firm employment on the lag of the change in firm's profits ( $\Delta\text{OIBD}_{t-1}$ ) plus the control variables. We can see from column (1) that profit growth is a significant determinant of employment growth, as predicted by our model. To test the imitationist hypothesis we created two dummy variables. The first, "Rival is Bigger and More Profitable" ( $\text{RB}_{t-1}$ ), takes on a value of 1 if the rival earns more and has more employees, 0 otherwise. The second, "Rival is Smaller and More Profitable" ( $\text{RS}_{t-1}$ ), takes on a value of 1 if the more profitable rival is smaller, 0 otherwise. Both variables are lagged one period. The results are presented in column (2). The dummy variables have the correct signs, in terms of the model's predictions, though just  $\text{RS}_{t-1}$  is statistically significant. This provides some support for the imitationist hypothesis: a smaller and more profitable rival will have a negative effect on growth. This result is not surprising. The choice of downsizing usually encounters less obstacles than increasing size (which may be prevented for example by capital or financial constraints).

Column (3) reports the test of the combined dynamics (Eq. 13), which confirms the results of the particular cases: both own profit and the dummies have the expected signs and are two of three coefficients are statistically significant. To conclude, our admittedly rough empirical exercise gives broad support to the rule-of-thumb dynamics that we proposed above.

## 9 Discussion and conclusion

This paper explored the long run size dynamics of firms-as-neural-networks playing a Cournot game in an uncertain environment. We derived a profit function for the firm that depends on its own size, its rival's size and environmental complexity. We then investigated long-run firm size dynamics resulting from "isolationist" and "imitationist" adaptive rules, which we compared to a benchmark "best response" case. First we found that when using simple adjustment rules, long run firm size is a function of initial firm size, initial rival's size, environmental complexity and the adjustment rate parameters. These variables interact in a non-linear way. We also found that under only

very precise initial conditions and parameter values do the firms converge to the Nash equilibrium size given by the best response. The reason is that the dynamics we consider tend to settle rapidly (no more than a few iterations) on a path that depends on initial conditions. Thus, our results suggest cautiousness in taking the Nash equilibrium as a focal point of simple dynamics (the standard “as if” argument).

Interestingly, we further find that when firms use simple adjustment rules, environmental complexity has a negative effect on size, even if the Nash equilibrium size is increasing in environmental complexity. This is particularly true in the isolationist case, and is explained by the negative correlation between profits and complexity: complex environments yield lower profits, and hence less incentives for firm growth. This paradoxical result disappears, though, when firms use the combined dynamics. Lastly we provide some empirical support of model using data from the Compustat database; the results are surprisingly supportive of our working hypotheses, given the roughness of the empirical exercise.

Our general conclusion is that in our model more efficient information processing and more complex adjustment rules would play a positive role in the long run profitability of the firm, and would deserve investment of resources. A possible extension of this paper would be to explicitly model costs and benefits of adjustment rules of varying efficiencies.

## Appendix

### Appendix A: Neural networks

This appendix briefly describes the working of the backward propagation network (BPN). For a more detailed treatment, the reader is referred to Skapura (1996). The BPN consists in a vector of  $\mathbf{x} \in \mathbb{R}^n$  inputs, and a collection of  $(m + o)$  processing nodes organized in layers. For our purposes we focus on a network with a single hidden layer. Inputs and nodes are connected by weights,  $\mathbf{w}^h \in \mathbb{R}^{n \times m}$ , which store the knowledge of the network ( $n$  is the number of elements in  $\mathbf{x}$ , and  $m$  is the number of agents in the network). The nodes are also connected, through the weight vector  $\mathbf{w}^o \in \mathbb{R}^{m \times o}$ , to an output vector  $\hat{\mathbf{y}} \in \mathbb{R}^o$ , where  $o$  is the number of outputs (2 in our case:  $\hat{\mathbf{y}} \equiv \{\hat{y}^i, \hat{q}_{-i}^i\}$ ).

The feed forward phase is given by

$$\hat{\mathbf{y}}_{1 \times o} = g \left[ g \left( \begin{matrix} \mathbf{x} & \cdot & \mathbf{w}^h \\ 1 \times n & & n \times m \end{matrix} \right) \mathbf{w}^o \right]_{m \times o},$$

where  $g(\cdot)$  is the sigmoid function ( $g(a) = 1/[1 + e^{-a}]$ ) that is applied both to the input to the hidden layer and to the output.

The error vector associated with the outputs of the network is  $\boldsymbol{\varepsilon} = \{(y_i - \hat{y}_i)^2\}$ ,  $i = 1, \dots, o$ , where  $y_i$  is the true value of the function, corresponding to the input vector  $\mathbf{x}$  (see Eq. 2). Total error is  $\xi = \sum_{i=1}^o \varepsilon_i$ . The learning

algorithm aims at minimizing the total error,  $\xi$ . The gradient of  $\xi$  with respect to the output-layer and the hidden layer weights respectively is

$$\frac{\partial \xi}{\partial \mathbf{w}^o} = -2 (\mathbf{y} - \hat{\mathbf{y}}) [\hat{\mathbf{y}}(\mathbf{1} - \hat{\mathbf{y}})] g (\mathbf{x} \cdot \mathbf{w}^h)$$

$$\frac{\partial \xi}{\partial \mathbf{w}^h} = -2 [(\mathbf{y} - \hat{\mathbf{y}}) [\hat{\mathbf{y}}(\mathbf{1} - \hat{\mathbf{y}})]] \mathbf{w}^o g' (\mathbf{x} \cdot \mathbf{w}^h) \mathbf{x},$$

since for the sigmoid function,  $\partial \hat{\mathbf{y}} / \partial \mathbf{w}^o = \hat{\mathbf{y}}(\mathbf{1} - \hat{\mathbf{y}})$ . The weights are then adjusted a small amount in the opposite (negative) direction of the gradient. We use a constant  $\eta$  to smooth the updating process. Define  $\delta^o = (\mathbf{y} - \hat{\mathbf{y}}) [\hat{\mathbf{y}}(\mathbf{1} - \hat{\mathbf{y}})] / 2$ . We then have the weight adjustment for the output layer as

$$\mathbf{w}^o(t + 1) = \mathbf{w}^o(t) + \eta \delta^o g (\mathbf{x} \cdot \mathbf{w}^h).$$

Similarly, for the hidden layer,  $\mathbf{w}^h(t + 1) = \mathbf{w}^h(t) + \eta \delta^h \mathbf{x}$ , where  $\delta^h = \mathbf{g}'(\mathbf{x} \cdot \mathbf{w}^h) \delta^o \mathbf{w}^o$ . When the updating of weights is finished, the firm views the next input pattern and repeats the weight-update process.

### Appendix B: Regression results for profit

Equation 11 is derived starting from the ANN model described in Appendix A. We built a data set by making random draws of  $n \in [5, 45]$ ,  $m_i \in [2, 20]$  and running the Cournot competition process for  $T = 250$  periods (we took care of random initial conditions by averaging over 25 runs). We recorded average profit for the two firms computed as in Eq. 3, and the values of  $m_1$ ,  $m_2$ , and  $n$ . This was repeated 10,000 times. We then ran a regression to obtain a precise polynomial form for profit. When dealing with a data set of randomly generated data, estimating the conditional expected value polynomial introduces a trade-off: too high a polynomial order runs the risk of “over-fitting” the data, while too low an order risks “under-fitting” the data.

**Table 5** Dep. var 10,000 ·  $\pi_1$ . Profit function for firm 1. Robust standard errors in parentheses. All coefficients stat. sig. at 99% or greater confidence level. Profit is multiplied by 10,000 to improve readability

Variable	Coefficient	Variable	Coefficient
<i>const</i>	270.6 (0.898)	$m_1 \cdot m_2$	-0.304 (0.004)
$m_1$	5.93 (0.229)	$n$	-2.201 (0.034)
$m_1^2$	-0.375 (0.023)	$n^2$	0.003 (0.001)
$m_1^3$	0.007 (0.001)	$m_1 \cdot m_2 \cdot n$	0.007 (0.00)
$m_2$	0.49 (0.056)	$m_2 \cdot n$	-0.016 (0.002)
<i>nobs.</i>	10,000		
$R^2$	0.864	<i>AIC</i>	7.183

In order to attempt to best fit the data two measures were used: the adjusted  $R^2$  and the Akaike information criterion (AIC). Our goal was to minimize the AIC, without lowering the adjusted  $R^2$  (Pindyck and Rubinfeld 1998). In addition, we only retained variables that had a greater than 99% level of confidence. Table 5 gives the complete results of the regression, which is reflected in Eq. 10.

## References

- Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev Econ Stud* 58:277–297
- Axtell R (2002) Non-cooperative dynamics of multi-agent teams. Brookings Institution CSED Working Paper, no. 27
- Barr J, Saraceno F (2002) A computational theory of the firm. *J Econ Behav Organ* 49:345–361
- Barr J, Saraceno F (2005) Cournot competition, organization and learning. *J Econ Dyn Control* 29(1–2):277–295
- Barr J, Saraceno F (2006) Firm size dynamics in a cournot competition model. In: Mathieu P, Beaufils B, Brandouy O (eds) *Artificial economics. Agent-based methods in finance, game theory and their applications Lecture notes in economics and mathematical systems*, vol 564. Springer Verlag, pp 65–76
- Barr J, Saraceno F (2008) Organization, learning and cooperation. *J Econ Behav Organ* (forthcoming)
- Chandler Jr, AD (1977) *The visible hand: the managerial revolution in American business*. Harvard Univ. Press, Boston
- Chang M-H, Harrington JE (2006) Agent-based models of organizations. In: Judd KL, Tesfatsion L (eds) *Handbook of computational economics*, vol 2, pp 273–1337
- Cyert RM, March JG (1963) *A behavioral theory of the firm*. Prentice-Hall, New Jersey
- DeCanio SJ, Watkins WE (1998) Information processing and organizational structure. *J Econ Behav Organ* 36:275–294
- Ericson R, Pakes A (1995) Markov perfect industry dynamics: a framework for empirical analysis. *Rev Econ Stud* 62(1):53–82
- Ernst B, Kieser A (1999) In search of explanations for the consulting explosion: a critical perspective on managers' decisions to contract a consultancy. Working Paper 99–87, Sonderforschungsbereich, University of Manheim
- Evans DS (1987) Tests of alternative theories of firm growth. *J Polit Econ* 95(4):657–674
- Ijiri Y, Simon HA (1977) *Skew distributions and the sizes of business firms*. North Holland, New York
- Jovanovic B (1982) Selection and the evolution of industry. *Econometrica* 50(3):649–70
- Kirby AJ (1988) Trade associations as information exchange mechanisms. *RAND J Econ* 19(1):138–146
- Lawrence P, Lorsch J (1986) *Organization and environment: managing differentiation and integration*, revised edition. Harvard Business School Press, Boston
- Mansfield E (1961) Technical change and the rate of imitation. *Econometrica* 29(4):741–766
- Mirvis PH (1997) *Human Resource management: leaders, laggards and followers*. Acad Manage Exec 11(2):43–56
- Nelson RR, Winter SG (1982) *An evolutionary theory of economic change*. Belknap Press of Harvard Univ. Press, Cambridge
- Nerlove M (1958) Adaptive expectations and the cobweb phenomena. *Q J Econ* 75:227–40
- Nickell SJ (1996) Competition and corporate performance. *J Polit Econ* 104(4):724–746
- Niyogi P (1998) *The informational complexity of learning*. Kluwer, Boston
- Novshek W, Sonnenschein H (1982) Fulfilled expectations Cournot duopoly with information acquisition and release. *Bell J Econ* 13:214–218
- Pindyck RS, Rubinfeld DL (1998) *Econometric models and economic forecasts*, 4th edn. McGraw-Hill, New York

- Radner R (1993) The organization of decentralized information processing. *Econometrica* 61:1109–1146
- Rivkin JW (2000) Imitation of complex strategies. *Manage Sci* 46(6):824–844
- Rivkin JW, Siggelkow N (2003) Balancing search and stability: interdependencies among elements of organizational design. *Manage Sci* 49(3):290–311
- Simon HA (1956) Rational choice and the structure of the environment. *Psychol Rev* 63(2): 129–138
- Simon HA (1976) *Administrative behavior*, 3rd edn. Free Press, New York
- Skapura DM (1996) *Building neural networks*. Addison-Wesley, New York
- Sutton J (1997) Gibrat's legacy. *J Econ Lit* XXXV:40–59
- Utterback JM (1994) *Mastering the dynamics of innovation*. Harvard Business School Press, Boston

Copyright of *Journal of Evolutionary Economics* is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Copyright of *Journal of Evolutionary Economics* is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.