AN ARTIFICIAL ECONOMY BASED ON REINFORCEMENT LEARNING AND AGENT BASED MODELING

Fernando Lozano[γ]
Jaime Lozano[α]
Mario García[ε]

First version: December 2004
This version: April 2007

In this paper, we employ techniques from artificial intelligence such as reinforcement learning and agent based modeling as building blocks of a computational model for an economy based on conventions. First we model the interaction among firms in the private sector. These firms behave in an information environment based on conventions, meaning that a firm is likely to behave as its neighbors if it observes that their actions lead to a good pay off. On the other hand, we propose the use of reinforcement learning as a computational model for the role of the government in the economy, as the agent that determines the fiscal policy, and whose objective is to maximize the growth of the economy. We present the implementation of a simulator of the proposed model based on SWARM, that employs the SARSA(λ) algorithm combined with a multilayer perceptron as the function approximation for the action value function.

Keywords: reinforcement learning; agent-based modeling; computational economics

1. Introduction

In the last decades there was a relative growth of the economic studies about individual learning. Rational Expectations literature changed the toolbox of economists and many macroeconomic and economic policy models changed as a consequence. New versions of IS-LM models based on rational expectations that showed Keynesian policy effects were generalized. Models of monetary policy now point out problems of dynamic inconsistency that are the direct outcome of the introduction of learning agents. Much of the economic literature about learning has been focused in the problem of explaining how individuals and organizations can learn in an environment that assumes that policies and government are given conditions. Usually the problem is to find the optimal policy for an economy in which individuals learn from government actions. Learning is understood as a Bayesian process of actualization of expectations that converges to the real model of the economy. These models do not specify the nature of the learning process beyond the problem of availability of information about the probable actions of the government. The agents act rationally in concordance with available information and a specific process of updating information. However, rational expectations models often have multiple equilibrium points, and there is not a consensus in the literature about how to explain why the economy reaches

[γ] Assistant Professor, Departament of Electrical Engineer, Universidad de los Andes, Carrera 1a, 18A-70, Bogotá, Colombia. Part of this work was done while the author was with the Department of Electronic Engineering, Universidad Javeriana, Bogotá, Colombia: flozano@uniandes.edu.co.
[α] Profesor, Economics Department Universidad del Rosario. Part of this work was done while the author was with the Department of Economics of Universidad Externado de Colombia, jlozano@urosario.edu.co
[ε] School of Economics, Universidad Nacional de Colombia,Bogotá, Colombia, mgarciamo@unal.edu.co

one of them and not the others. When thinking about the role of the government in economic growth, some economists consider that it has a coordinating role, and that it pushes the economy toward a more desirable equilibrium point. The later assumes that there is a great normative clarity about what is desirable for society and about what are the institutional capacities the government has to develop what it proposes as a target. The recognition of the existence of complementarities between government and private actions has permitted to broaden macroeconomic discussion about coordination failures of the economic system and the scope of economic policy and to reformulate the analysis of the government's institutional ability to push economic growth with its fiscal policy. In fact the existence of an order of multiple equilibriums depends on the optimism of private agents, and can explain the possibility that the government action can complement private actions. Strategic complementarities are those in which the increase of the effort of other agents conduces causes an agent to increase its own. This relationship between both efforts is the base of multiple equilibriums and strategic complementarities can generate multipliers effects[1]. All this depends on the degree to which government can agree with private interests, and its skills to act first and sustain its actions.

In these models the effectiveness of a policy depends on the temporal ordering of the actions the government and the private sector follow. However there is no definitive theoretical argument that supports the idea that the government cannot learn to act first and to have a more effective fiscal policy that leads to better rates of long run economic growth[2]. But to think about this possibility, understood as an emergence process[3], requires an approximation to some open theoretical questions and figuring out appropriate formal tools that allow us to show how the interaction of the government and the private sector can maximize economic growth.

In this paper, we take a different path. We employ techniques from artificial intelligence such as reinforcement learning and agent based modeling as building blocks of a computational model for an economy that permits us to shed light into the problems mentioned above. First we use the agent based capabilities of SWARM[4] to model the interaction among firms in the private sector. These firms behave in an information environment based on conventions, meaning that a firm is likely to behave as its neighbors if it observes that their actions lead to a good pay off. On the other hand, we propose the use of reinforcement learning as a computational model for the role of the government in

---

[1] Russell Cooper. Coordination games: complementarities and macroeconomics. Cambridge University Press, 1999.

[2] Cf. Thomas Sargent. Optimal fiscal policy in a linear stochastic economy. Technical report, Hoover Institution, University of Chicago, 1998 and Evans George W; Seppo Honkapohja ,2001.

[3] An emergent property may be defined as a feature of a complex system that:
a) Can be described in terms of macro-or-aggregate level concepts, without reference to the attributes of specific micro-level entities
b) Persists for time periods significantly greater than those required for describing the underlying microinteractions, and
c) Is not explicable entirely in terms of the microporperties of elemental components of the system.
See Lane 1993a, p.93.

[4] SWARM. Software package for multiagent simulations of complex systems. Swarm Development Group:http://wiki.swarm.org.

the economy, as the agent that determines the fiscal policy, and whose objective is to maximize the growth of the economy.

In the next section, we explain how the interaction of the government with the economy can be cast in the reinforcement learning formalism and explain the algorithms employed in our simulator. Next we show some initial results of the application of reinforcement learning to the maximization of economic growth by means of the fiscal policy of the government.

2. Reinforcement Learning and a Government that learns

According to reinforcement learning theory[5], the learning agent is immersed in an environment that is usually complex and uncertain. The agent is supposed to learn to reach a goal by means of a series of actions. The agent observes the state of the environment and chooses and executes an action. The action in turn, has consequences that modify the environment. The agent does not receive information regarding the intrinsic value of each of the actions it executes, but only partial information, in the form of a reward signal (or reinforcement signal). The objective of the learning process is to find an optimal policy that tells the agent how to select an action in a given state of the environment, in such a way that the total reward is maximized (which usually means that the agent reaches its goal at minimum cost).

A fundamental issue in reinforcement learning algorithms is the balance between exploration and exploitation. If an agent can find an action that yields a high reward, it can decide to stick to that action in the future (exploitation). However, that action may be not optimal in the long run, so it may be better to try a different action on occasion (exploration). In economics, models based on reinforcement learning have been applied to microeconomic problems[6] but applications to macroeconomic situations are scarce. In the macroeconomic model that we propose, the learning activity of the government can be represented with the reinforcement learning paradigm. The objective of the government as a learning agent is to learn an optimal policy that maximizes economic growth. By means of its fiscal policy, the government modifies the economic environment, without knowing,

---

[5] Richard S. Sutton and Andrew G. Barto. ***Reinforcement Learning***. MIT Press, 1998.

[6] For a complete analysis of reinforcement learning theory in agent based models see Duffy, J. (2006). "Agent-based models and human-subject experiments", and Brenner, T. (2006). "Agent learning representation: Advice on modelling economic learning" in Testfasion L and Judd K. eds. ***Handbook of Computational Economics*** vol.2, Elsevier. See too Borgers Tilman and Rajiv Sarin. Learning through reinforcement and replicator dynamics. ***Journal of Economic Theory***, 77:1–14, 1997, Ed Hopkins. Learning, matching and aggregation. ***Games and Economic Behavior***, 26:79–110, 1999. Ido Erev and Alvin Roth. Predicting how people play games: reinforcement learning in experimental games with unique mixed strategy equilibria. ***American Economic Review***, 88:848–881, 1998. Ido Erev, Yoella Bereby-Meyer, and Alvin Roth. The effect of adding a constant to all payoffs: experimental investigation, and implications for reinforcement learning models. ***Journal of Economic Behavior and Organization***, 39:111–128, 1999. Nick Feltovich. Equilibrium and reinforcement learning in private information games: an experimental study. ***Journal of Economic Dynamics and Control***, 23:1605–1632, 1999.

in principle, a model of the economy. The government needs to learn to identify actions that are optimal in the lung run, in the sense that they maximize the growth of a decentralized economy.

For simplicity, we assume that the government interacts with the environment consisting of a set of firms, in a sequence of discrete steps t = 0, 1, 2, . . At a given time t, the government receives a representation of the state of the environment $s_t \in S$, where S is the set of all possible states of the environment. Based on s, the agent selects an action $a_t \in A(s_t)$ where $A(s_t)$ is the set of possible actions at state s_t. By executing the action, the agent receives a reward R_{t+1} and in turn, modifies the environment that goes to a new state s_{t+1}.We assume that the reinforcement task at hand can be modeled as a Markov decision process.

A policy P is a mapping S ×A $\rightarrow$[0, 1] that assigns a probability to each action $a_t \in A(s_t)$ of being executed when the state of the environment is s_t. The task of the reinforcement learning algorithm is to find a policy that maximizes the expected value of the return received by the agent. This return is defined as the discounted sum of the rewards received over the future:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

where $\gamma \in$ [0, 1] is the discount rate.

The reinforcement learning algorithm with an on-policy strategy iterates to basic steps.4 In the first step, it computes an estimate of the action value function (initially, for an arbitrary policy):

$$Q(s,a) = \mathbb{E}\{R_t | s_t = s, a_t = a\}$$

which gives the expected value of the return when action a is taken in state s. In the second step, the policy is improved by selecting the next action in a greedy fashion according to the current estimate of the function Q. To maintain exploration, an arbitrary action is selected with a small probability ε.

In the SARSA(λ) algorithm4 (which we implement in our simulator), the return value of an state action pair is approximated as $r_{t+1} + \lambda Q_t(s_{t+1,} a_{t+1})$, where s_{t+1} is the resulting state after executing action a_t at, and Q_t is the current estimate of the action value function. The algorithm updates all state actions pairs in the following way:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha \delta_t e_t(s,a)$$

where

$$\delta_t = r_{t+1} + \lambda Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)$$

and $e_t(s,a)$ is the eligibility trace at (s,a) given by:

$$e_t(s,a) = \begin{cases} \gamma\lambda e_{t-1}(s,a) + 1 \text{ if } s = s_t \text{ anf } a = a_t; \\ \gamma\lambda e_{t-1}(s,a) \qquad \text{otherwise} \end{cases}$$

3. An artificial economy

We have written a simulator of an economy based on conventions where the government uses reinforcement learning to search for a good expenditure policy. The simulator was written using the Swarm library for agent based modeling. Figure 1 shows the general structure of the economy as well as the flow of information between the different components.

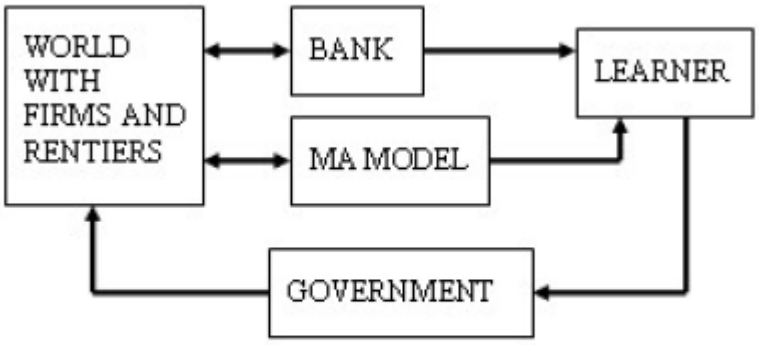

Fig. 1. General blocks diagram of the economy. Arrows depict the flow of information

Firms live in a two dimensional world where the propagation of expectations occur. A firm looks for a more or less optimistic rule (which determines how much the firm invests) depending on the value of the utilization of capacity (set by MA model) and the increase or decrease of the national income. For example, if the national income decreases an the utilization of capacity is low, the firm looks for a more optimistic rule among its neighbors. These changes occur only if the utilization of capacity has remained unchanged for a number of periods, and the firm has reason to believe that the situation is permanent. We allow the rule of each firm to be changed with a small probability to a random value. This allows us to maintain diversity in the pool of firms.

Each firms maintains sight deposits, time deposits, and debt. According to the interest rate determined by the bank, it computes the amount of interest due to the bank and pays the interests with the sight deposits or, if they are not enough, with money borrowed from the

bank. In the later case, the firm sets a flag indicating that it has entered a Ponzi situation. Similarly, the firm determines how much to invest (according to the rule set previously) and uses the sight deposits and/or increases its debt. Any leftover cash is used to pay debt or put in time deposits. When its capital depreciates completely, the firm becomes a rentier. The multiplier accelerator (MA) model collects information from all the firms and rentiers and implements the dynamics of the economy, according to the equation:

$$Y = \frac{I - D + E}{(1 - (1 - s)(1 - t))}$$

where Y, I,D and E denote national income, investment, depreciation and government expenditure respectively, s is the saving rate and t the tax rate. The model then compares the ratio between capital and national income with a set of thresholds and sets the variable of utilization of capacity accordingly. The bank determines the interest rate as a linear function of the change in the number of firms and rentiers and the change in the national income. It also eliminates firms and rentiers that become bankrupt.

The learner uses the gradient descent version of the SARSA(λ) algorithm to learn a good policy for the determination of the government expenditure. The information given to the learner consist of:

- Past history of the national income.
- Past history of the total debt.
- Past history of the number of firms and rentiers, including bankrupts.

The learner returns an action in the form of a percentage change in the expenditure in a given range and with a given resolution within that range. As mentioned previously, the SARSA(λ) algorithm searches for an optimal policy, by estimating the State-action value function Q, and selecting greedily the next action according to the current estimate of Q at the current state. Since the state space is continuous valued, it is necessary to use a function approximation technique. We use a multilayer perceptron with one hidden layer to model Q. Thus, if w is the vector of weights of the MLP, the basic update equation is:

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \alpha \delta_t \mathbf{e}_t$$

where

$$\delta_t = r_{t+1} + \lambda Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)$$

and *e* is the vector of eligibility traces computed as:

$$\mathbf{e}_t = \gamma \lambda \mathbf{e}_{t-1} + \nabla_{\mathbf{w}} Q_t(s_t, a_t)$$

The gradient $\nabla_w Q_t(s_t, a_t)$ is computed using the back-propagation algorithm. The reward signal is designed to encourage an increase in the national income while maintaining the economy in good shape. This implies following rules that maintain debt capacity of government under control and that the firms don't dead. We compute *r* as a linear function of the percentage change in the national income, the total debt, and the change in the number of firms. When the number of firms goes bellow a given threshold, the reward is a large negative number.

4. Simulation

We present results of the use of our simulator with a small economy which initially has 100 firms. In figure 2 we present the evolution of the first few steps of the simulation when learning has not occurred. We can observe in this graphics the cyclical nature of the economy, for example in the behavior of the national income. In figure 3 we show a snapshot of the world were the firms live and interact. Colors indicate the rule currently in use by the firm at that location.

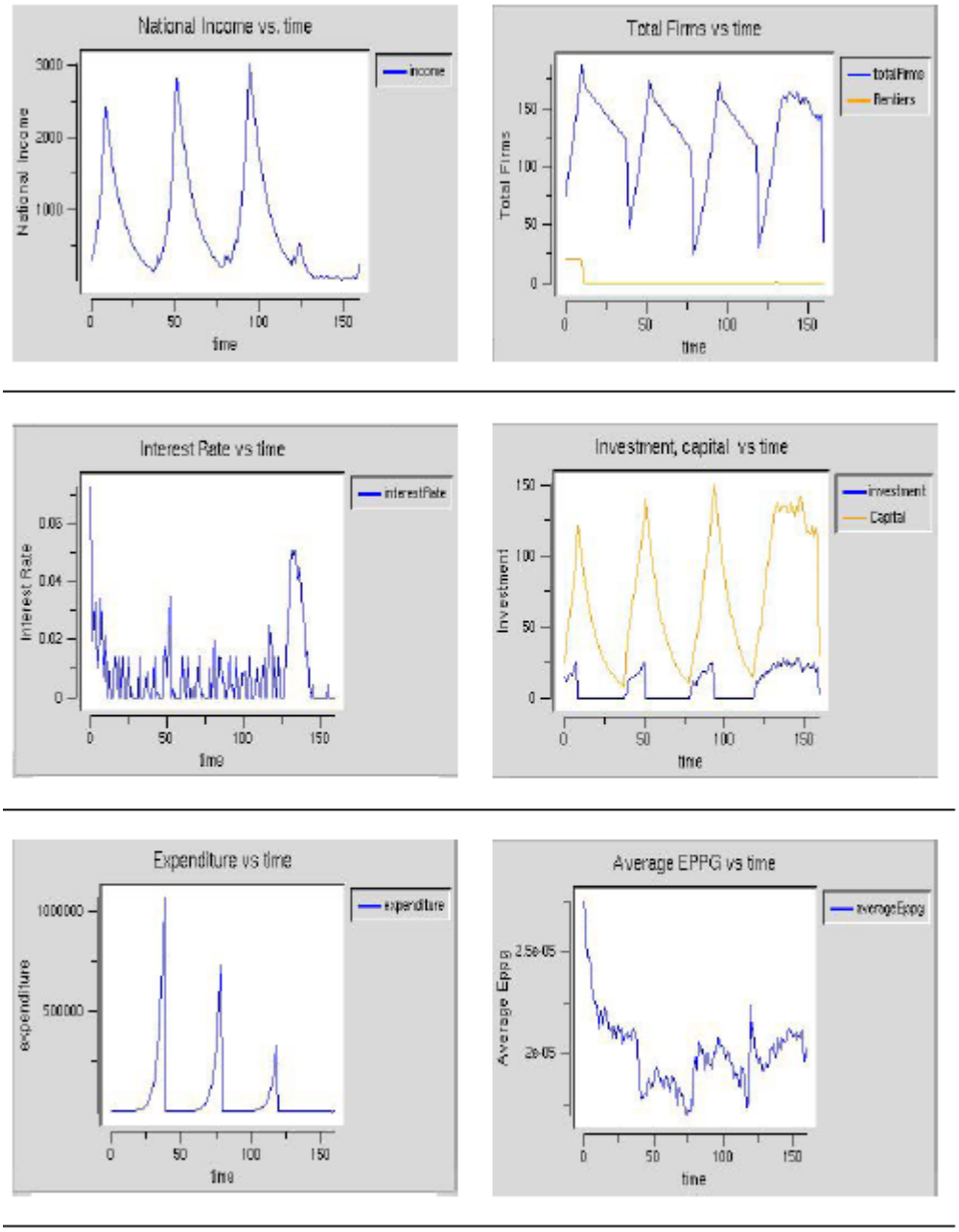

Fig. 2. Simulation of the artificial economy with conventions

This simulation implements a learner that uses as input information the past three values of the national income, debt, and number of firms in the world. We have used 8 neurons in the hidden layer of the multilayer perceptron and set $\gamma = 0.8$, $\lambda = 0.8$ and a learning rate of 0.0001.

Fig. 3. A snapshot of the world were the firms live and interact. Colors indicate the rule currently in use by the firm at that location.

In figure 4 we show the behavior of the national income after a few thousand iterations. We can see that although the cyclic behavior of the economy continues and the government is not able to prevent sudden collapses of the economy, these collapses are less frequent and the peak value of the income has an increasing trend.

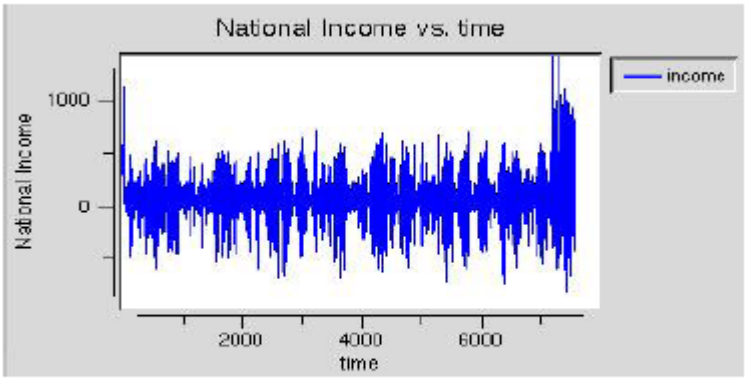

Fig. 4. Behavior of national income after learning

5. Conclusions

We have used reinforcement learning and agent based modeling as building blocks of a computational model for an economy based on conventions. We have pointed out some dimensions of the problem of modeling government learning in an economy in which entrepreneurs expectations determine the long run tendencies of economic growth: coordination of expectations with conventions define the limits of the economic growth, and these limits are related to the rules that the government uses to learn. Further research

will include improvements in the convergence of the reinforcement learning algorithm, in particular in the function approximation process.

Acknowledgements
This work was funded by Colciencias, the Inter American Development Bank and the Observatorio de Ciencia y Tecnología OCyT.

References

1. Thomas Sargent. *Optimal fiscal policy in a linear stochastic economy*. Technical report, Hoover Institution, University of Chicago, 1998.
2. Lane, D. A. Artificial Worlds and Economics, Part I", *Journal of Evolutionary Economics*, 3: 89{107), 1993a.; Lane, D. A. Artificial Worlds and Economics, Part II", *Journal of Evolutionary Economics*, 3: 177{97), 1993.
3. Evans George W; Seppo Honkapohja, *Learning and Expectations in Macroeconomics*, Princeton University Press, 2001.
4. Russell Cooper. *Coordination games: complementarities and macroeconomics*. Cambridge University Press, 1999.
5. SWARM. Software package for multiagent simulations of complex systems. Swarm Development Group:http://wiki.swarm.org.
6. Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning*. MIT Press, 1998.
7. Borgers Tilman and Rajiv Sarin. Learning thorugh reinforcement and replicator dynamics. *Journal of Economic Theory*, 77:1–14, 1997.
8. Ed Hopkins. Learning, mathing and aggregation. *Games and economic behavior*, 26:79–110, 1999.
9. Ido Erev and Alvin Roth. Predicting how people play games: reinforcement learning in experimental games with unique mixed strategy equilibria. *American EconomicReview*, 88:848–881, 1998.
10. Ido Erev, Yoella Bereby-Meyer, and Alvin Roth. The effect of adding a constant to all payoffs: experimental investigation, and implications for reinforcement learning models. *Journal of Economic Behavior and Organization*, 39:111–128, 1999.
11. Nick Feltovich. Equilibrium and reinforcement learning in private information games: an experimental study. *Journal of Economic Dynamics and Control*, 23:1605–1632, 1999.
12. Duffy, J."Agent-based models and human-subject experiments", Testfasion L and Judd K. eds. *Handbook of Computational Economics* vol.2, Elsevier, 2006.
13. Brenner, T. "Agent Learning Representation: Advice on Modelling Economic Learning" in Testfasion L and Judd K. eds. *Handbook of Computational Economics* vol.2, Elsevier, 2006.