Modeling an Economy with Policy Gradients

Nicholas Hirning
Stanford University
nhirning@stanford.edu

Rory Lipkis
Stanford University
rlipkis@stanford.edu

Andrew Chen
Stanford University
asjchen@stanford.edu

Abstract

TODO

I. Introduction

In simple models of market economies, a common assumption is that each citizen will act in their own best interest. This selfish behavior plays a major role in determining the efficiency of such markets and leads to many of the phenomena predicted and analyzed by economists. The advantage of this assumption is that once a utility function is chosen for these models, each agent in the economy acts simply to maximize their future utility. In practice, this lends itself naturally to construction via a multi-agent reinforcement learning model.

A simple model taught in introductory economics classes is the circular flow model. In this setup, there are two types of agents in the economy: firms (or businesses), and people (or households). These agents interact in two different markets, the goods market and the labor market. In the goods market, people pay currency to firms in exchange for goods. In the labor market, firms pay people for their labor to produce more goods. This basic model is diagrammed in figure 1. The name comes from the fact that currency flows clockwise in a circle between firms and people.

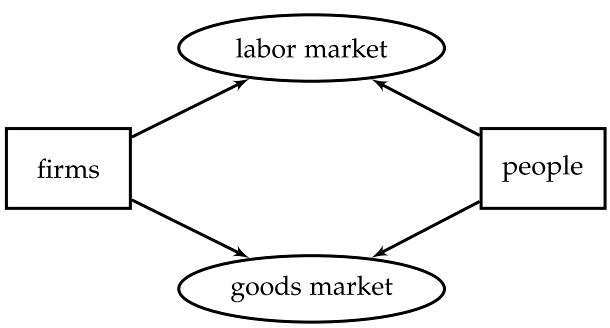

Figure 1: *Basic schematic of the circular flow model, a simple economic model in which society is composed of two types of agents (firms and people) that interact in two distinct markets.*

While the circular flow model does not include many important economic factors, it has the benefit of being simple enough to serve as a starting point for high-level analysis. For more in-depth economic phenomena, one would likely benefit from including more markets and agents; for example, the five-sector economy model includes government, financial markets, and overseas influences. However, as we aim to provide a proof-of-concept for modeling economies with multi-agent reinforcement learning, the circular flow model serves as a simple foundation that can potentially produce nontrivial dynamics.

In the following work, we explore how various economic phenomena can be simulated and analyzed by applying reinforcement learning to the circular flow model. While we are limited by the simplicity of the model, there are many economic metrics that can be directly investigated in such a model. For example, the gross domestic product (GDP) is widely touted as a comprehensive measure of a country's economic health. The GDP is defined by the following equation:

$$\text{GDP} = C + G + I + N,$$

where C stands for consumer spending, G for government spending, I for private domestic investments, and N for net exports. In practice, the vast majority of GDP is contained in consumer spending (accounting for roughly 2/3 of the U.S. GDP). While the circular flow model cannot account for net exports or government spending, the total consumer spending and private domestic investments can be approximated by the total quantity of money exchanged in the goods and labor markets. Thus, such a model can lead to investigation of how GDP changes over time and as a function of the number of agents of each type.

Another important economic focus is on income and wealth inequality, both of which are manifested in the circular flow model. A common approach to measuring inequality is to use the Gini coefficient. The Gini coefficient of a distribution p with mean μ is given by

$$G = \frac{1}{2\mu} \iint p(x)p(y)|x - y| \, \mathrm{d}x \, \mathrm{d}y.$$

In a discrete scenario, this just becomes the average absolute difference of all pairs of items (e.g., income or absolute wealth) divided by twice the mean. Given its relevance to modern society, understanding how income inequality develops and progresses over time and as a function of initial wealth and skill distribution is quite important. We aim to perform basic analyses of both of these phenomena in the work presented herein.

II. Literature Review

Despite the applicability of reinforcement learning techniques to economics, there are relatively few examples of such applications in existing literature. One particularly relevant example is a paper by Lozano et. al. [1] which focuses on modeling an economy "based on conventions." In this work, the authors apply the SARSA(λ) algorithm in a government agent to search for a good expenditure policy. In this case, the authors avoid applying complicated multi-agent reinforcement learning by modeling each firm with a simple algorithm in which firms can copy neighboring firms or (with some low probability) switch to a random policy. The government attempts to increase national income and decrease debt based on some limited knowledge of the existing firms and historical records of income, debt, etc.

One important issue that the authors of [1] grappled with was the vast (and continuous) state space. In this case, they applied a multilayer-perceptron (MLP) with a single hidden layer to approximate the Q-function for SARSA(λ). Given the vast array of actions and states within any economy, approximation with an MLP model or similar is common in such scenarios. The authors found limited success with their model, with the economy collapsing less frequently over time due to the government's policies.

Another relevant work by Tesauro & Kephart involved investigation of the behavior of reinforcement learning agents in the presence of other adaptive agents for economic purposes [2]. In particular, Tesauro & Kephart consider two selling agents competing in an economic model where each agent applied Q-learning. In this case, the state and action space was small enough that the Q-learning could proceed directly. Furthermore, the consumers in the model act via a simple, greedy rule as opposed to acting as an independent, market-shaping force. Another difference between Tesauro & Kephart and the model proposed here is that the former model does not assume simultaneous price-setting; instead, the two selling agents take turns adjusting prices in response to each other. This turn-based approach is slightly less realistic than the simultaneous setting, but it allowed the authors to frame the problem as a two-player, alternating-turn, arbitrary-sum Markov game. Additionally, the authors spent significant amount of their time computing the Q-function for one of the sellers when the other seller was assumed myopic; they investigate into simultaneous Q-learning as well, but note that no convergence proofs exist for such cases.

TODO: Moar.

III. Methods

IV. Results and Data Analysis

V. Conclusion

VI. Acknowledgments

The authors feel that they all contributed equally to the work presented. In particular, Andrew Chen implemented the actor-critic reinforcement learning algorithm, Nicholas Hirning implemented the circular flow model architecture, and Rory Lipkis performed hyper-parameter optimization and data analysis. All three authors collaborated on implementing and debugging the REINFORCE policy gradient algorithm and other areas of the code.

The authors would like to thank the CS 238 staff for a really great time throughout this course `;)`.

References

[1] Fernando Lozano, Jaime Lozano, and Mario García. An artificial economy based on reinforcement learning and agent based modeling. Documentos de Trabajo 003907, Universidad del Rosario, April 2007.

[2] Gerald Tesauro and Jeffrey Kephart. Pricing in agent economies using multi-agent q-learning. *Autonomous Agent and Multi-Agent Systems*, 5, 10 1999.