Project Milestone 2: Finding Optimal Taxation Policy With RL

Ardhito Nurhadyansah[1] and Ian Suryadi Timothy H[2]

[1]2106750206
[2]2106750875

Abstract

In this project, we simulate the economic output, given a tax policy. This project is adapted from the work of Zheng et al. (2020), where the model is comprises of two interconnected Markov Decision Process. The inner-loop MDP is the agents MDP, which have partial observations and interact to each other in a gather-and-build simulation. The outer-loop MDP models a social planner that find the optimal tax policy to maximize a social welfare function, which is formulated by a product of equality and economic productivity.

1 Problem Definition

Tax is one tool for a country to reduce inequality. Traditional tax systems are often derived from static models or rely on simplifying assumptions. However, real-world economies are dynamic and complex, featuring uncertain behaviors, heterogeneous agents, and evolving macroeconomic conditions. In this project, we are experimenting with tax policies and observe its effect on equality and productivity. This project is adapted from the work of Zheng et al. (2020), with some adjustments regarding compute power availability. The problem is formulated as two interconnected Markov Decision Processes (MDPs):

- Agents' MDP (inner loop): models the individual behaviors of economic agents.
- Planner's MDP (outer loop): models the tax-setting problem to optimize a social welfare function.

1.1 Agents' MDP (Inner Loop)

Each agent operates in a partially observable environment defined by:

- **State Space (S):**
  - Local Spatial Information: Agent's position in a 2D grid and nearby environmental features (e.g., resource tiles, obstacles).
  - Resource Availability: Status of resource regeneration (e.g., wood and stone).
  - Agent Attributes: Individual coin endowment, resource inventory, accumulated labor, and skill level (which affects income generation, e.g., coins earned per house built).
  - Market State: Public trading offers (bids and asks).
  - Tax Policy: The current tax schedule applied to income.
- **Action Space (A):**

$$a \in \{\text{Movement, Resource Collection, Building, Trading}\}$$

- **Transition Function (T):** The dynamics update the state based on:

  - Movements and interactions (e.g., collisions, blocked paths).
  - Stochastic resource regeneration.
  - Market execution and trade outcomes.
  - Periodic taxation that applies a bracketed tax schedule and redistributes income.
- **Reward Function (r):** The instantaneous reward for agent i at time t is defined as the change in its utility:

$$r_{i,t} = u(x_{i,t}, l_{i,t}) - u(x_{i,t-1}, l_{i,t-1}),$$

where the utility function is defined as

$$u(x, l) = \frac{x^{1-\eta} - 1}{1 - \eta} - l, \quad \eta > 0,$$

with x representing coins and l the cumulative labor.
- **Discount Factor (γ):** Each agent maximizes its expected discounted return:

$$\max_{\pi_i} \mathbb{E}\left[\sum_{t=0}^{H} \gamma^t \, r_{i,t}\right].$$

1.2 Planner's MDP (Outer Loop)

The planner sets tax policies with the goal of maximizing overall social welfare, based on aggregated economic outcomes.

- **State Space (S_p):** The planner observes aggregated indicators:
  - Aggregate Wealth Distribution: Summary statistics (e.g., mean, variance, Gini coefficient) of agents' incomes.
  - Productivity: Total income generated by the agents.
  - Market and Resource Summaries: Global resource availability and trading volumes.
  - Tax History: Current and previous tax schedules and tax revenue collections.
- **Action Space (A_p):** The planner's decision is to set a tax schedule:

$$a_p = [\tau_0, \tau_1, \ldots, \tau_{B-1}],$$

where each $\tau_b \in [0, 1]$ is the marginal tax rate for bracket b.
- **Transition Function (T_p):** The state evolves based on the effect of the tax policy on the agents' behaviors during a tax period. The agents' adaptive responses determine the new aggregate economic state.
- **Reward Function (r_p):** The planner receives a reward based on the improvement in a social welfare function. For example:

$$r_p = \text{swf}(s'_p) - \text{swf}(s_p),$$

where a candidate social welfare function is:

$$\text{swf} = \text{Equality}(x_c) \times \text{Productivity}(x_c),$$

or alternatively a weighted sum of individual utilities:

$$\text{swf} = \sum_i \omega_i \, u(x_{c,i}, l_i).$$

- **Discount Factor (γ_p):** The planner maximizes:

$$\max_{\pi_p} \mathbb{E}\left[\sum_{p=0}^{P} \gamma_p^p \, r_p\right].$$

2 Solution Design

A two-level deep reinforcement learning framework will be used, integrating the agents and planner's MDP.

2.1 Inner Loop (Agents' Learning)

- Agents interact in the environment and learn policies π_i using deep RL algorithms (e.g., PPO).
- Their objective is to maximize the expected discounted return based on their MDP, with rewards that reflect changes in utility (coins versus labor cost).

2.2 Outer Loop (Planner's Learning)

- The planner observes an aggregate state s_p (wealth distribution, productivity, etc.) and sets a tax schedule a_p for the upcoming tax period.
- Its policy π_p is trained with deep RL methods to maximize improvements in social welfare:

$$r_p = \text{swf}(s'_p) - \text{swf}(s_p).$$

2.3 Interaction Dynamics

- Influence of Planner on Agents: The tax schedule (action a_p) directly affects agents' rewards by modifying post-tax incomes.
- Influence of Agents on Planner: The agents' adaptive responses under the new tax policy determine the next aggregate state s'_p that the planner observes.

3 Preliminary Analysis

Our preliminary analysis builds upon the findings reported by Zheng et al. (2020), that the two-level reinforcement learning framework can capture complex dynamics. Specifically, the following observations were made:

- **Agents' Adaptive Behavior:**
  - Agents naturally exhibit specialization, where lower-skilled agents tend to focus on resource collection while higher-skilled agents emphasize building. This division of labor is a key emergent property observed in the simulation.
  - Agents dynamically adjust their strategies in response to the imposed tax schedule, demonstrating the non-stationarity of the inner-loop MDP and confirming that individual behaviors are influenced by external economic incentives.
- **Impact of Tax Policies on Social Welfare:**
  - The RL-based planner is capable of setting tax policies that strike an effective balance between income equality and overall productivity. They report that the AI-driven tax policy achieves approximately a 16% improvement in the product of equality and productivity relative to baseline methods.
  - The simulation reveals that agents sometimes engage in tax-gaming strategies, such as alternating between high and low income periods, in order to reduce their effective tax burdens.
- **Convergence and Stability:**
  - Curriculum learning and entropy regularization was crucial in stabilizing the training process. These techniques help both the agents and the planner converge to stable policies despite the inherent non-stationarity.
  - The planner's ability to adjust tax policies based on aggregated economic indicators (e.g., wealth distribution, overall productivity) ensures that the outer-loop MDP effectively guides the inner-loop dynamics towards improved social welfare outcomes.

Building on the concepts given in the work of Zheng et al. (2020), our project aims to adapt the methodology for doing economic simulations, with some modifications on the training process regarding compute power availability.

4 Citations and References

References

Zheng, S., Trott, A., Srinivasa, S., Naik, N., Gruesbeck, M., Parkes, D. C., and Socher, R. (2020). The ai economist: Improving equality and productivity with ai-driven tax policies. (pages 1 and 3)