Quadratic Menu Costs with Reinforcement Learning

Trevor Gallen

Econ 64200

Fall 2023

This homework asks you to write down and run a simple dynamic problem using reinforcement learning. **Deliverables**

- You should have a word/LaTeXdocument that has three sections:
  1. Discusses the model and answers the questions I pose throughout.
  2. Contains the tables and figures you will produce.
  3. Contains a discussion of your programming choices if you had to make any.
- You should have a Matlab file or set of files (zipped) that contain **all** your programs and raw data. There should be a file called "Main.M" that produces everything I need in one click.

1 Model

Each period, firms begin the period with prices p_{t-1} and observe optimal price $\overline{P}_t$. They may either change their price p_t, which costs $\phi(p_t - p_{t-1})^2$, or leave it fixed at p_{t-1}. Their value function is:

$$V(p_{t-1}, \overline{P}_t) = \max_{p_t} -(p_t - \overline{P}_t)^2 - \phi(p_t - p_{t-1})^2 + \beta V(p_t, \overline{P}_{t+1})$$

Where:

$$\overline{P}_t = \rho \overline{P}_{t-1} + \epsilon \quad \epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$$

Let $\rho = 0.95$, $\sigma = 0.01$, and $\phi = 0.1$. Discount factor is 0.95.

2 Problems

Question 1 Do the following in matlab:

- Define an environment for the problem: observation (2x1) and action info (1x1).
- Write out a reset function for state variables p_{t-1} and $\overline{P}_t$
- Write out a step function that takes in $\overline{P}_t$, p_{t-1} and p_t and generates the reward and next $\overline{P}_{t+1}$
- Define a critic network, and actor network, and an agent
- Train the agent
- Display the agent's optimal policy function for p_t as a function of $\overline{P}_t$ and p_{t-1}.