Optimal Monetary Policy using Reinforcement Learning

Project Overview

Central Bank learns its optimal reaction function for the nominal interest rate, interacting with the proposed economic environment that evolves with the respective transition equations.

Central Bank, as the agent, observes the current state of the economy but is not informed about the transition equations. It interacts with the unknown environment and observes inflation rate (π_t) and output gap (y_t) for each time period. Given the observations, the agent takes the action, setting the nominal interest rate (i_t), and receives a reward. Thus, we experiment the economy's reaction to the monetary policy actions.

The reward function (r_t) of the agent depends on the deviations from the fulfilment of the inflation and output target, defined as the following:

$$r_t(x_t^z, i_t) = -\omega_\pi(\pi_{t+1} - \pi^*)^2 - \omega_y y_{t+1}^2$$

where:

$$\omega_\pi = \omega_y = 0.5, \quad \pi^* = 2\%.$$

The agent follows a dynamic programming setting, optimising its objective function for each time period. Bellman equation is defined for each time stamp to reflect the value function of the agent with each policy iteration. Given the reward function at the end of the time period, the agent evaluates the past behaviour and adapts its action accordingly. Hence, the learning takes place through (r_t).

Iteration of this model for T times ultimately yields the optimal Policy Reaction Function (PRF).

Appendix

One key aspect of implementation of Reinforcement Learning to monetary policy-making is the mitigation of "model uncertainty" problem, given that the algorithm inherently considers Zero Lower Bound, convex Phillips Curve and asymmetric preferences. The major considerations include data uncertainty, shock uncertainty, parameter uncertainty and Knightian uncertainty that yields conservative (*Brainard Conservatism*) actions by the agent to not to take excessive risk, preventing the freedom of decision-making. This is thanks to the specification of the Reinforcement Learning algorithm which does not require perfect information in the environment.

As a reference to our model, the Loss Function is generally defined under the following setting:

Consider the following economy. The demand for output is given by the dynamic IS relationship:

$$x_t = E_t[x_{t+1}] - (i_t - E_t[\pi_{t+1}] - \delta)/\sigma + v_t$$

where x_t is the output gap, i_t is the policy rate, π_t is inflation and v_t is an unpredictable demand shock which is observed at the start of period *t*. The supply of output is given by the New Keynesian Phillips curve:

$$\pi_t = E_t[\pi_{t+1}] + x_t + u_t$$

where u_t is an unpredictable cost shock which is also observed at the start of period *t* and has a zero mean and variance s^2, i.e. $E_t[u_{t+1}] = 0$ and $E_t[u_{t+1}^2] = s^2$. Finally, the policy maker's within-period loss function is:

$$L_t = \frac{1}{2}[\pi_t^2 + x_t^2]$$