**Kathmandu University School of Management**

**Balkumari, Lalitpur**

****

**Optimal Monetary Policy using Q-Learning and Markov Decision Process.**

**Submitted by:**

Nimesh Mani Risal (23916)

**Submitted to:**

**Dipesh Karki**

Asst. Professor, KUSOM

6th Feb, 2025

**Introduction**

Economic modelling is one of the very crucial steps in policy-making, allowing central banks and financial institutions to simulate various policy interventions before implementing them. One of the primary challenges in economic modelling is finding the optimal policy that minimizes loss and maximizes profitability of a firm, industry and a country. Having said that, in this paper the primary focus is to simulate an optimal monetary policy by reducing the output gap and maintaining the optimal inflation rate. In this process an AI model called reinforcement learning, particularly Q-learning, has been used to develop an optimized monetary policy framework. This approach has been enhanced by integrating Markov Decision Process (MDP)-based state transitions, which help guide learning in a structured and probabilistic manner.

The main objective is to use machine learning techniques to simulate and optimize economic decision-making, specifically in adjusting interest rates to minimize the output gap and stabilize inflation. The algorithm learns by interacting with an economic environment, where different monetary policy actions lead to distinct outcomes. Over multiple simulations, the agent discovers a policy that results in the most stable economic conditions.

To validate the effectiveness of the learned policy, multiple economic simulations has been conducted, tracking how states evolve over time. Further visualization of economic state transitions using network graphs has been done, allowing for an intuitive understanding of the trajectory an economy follows under different policy actions.

**Data Preprocessing and Discretization**

The dataset used in this study comprises key economic indicators such as:

* Gross Domestic Product (GDP)
* Inflation Rate
* Money Supply (M2)
* Currency Circulation Ratio

These variables serve as inputs for the reinforcement learning model used, defining the state space in the MDP framework. However, working with continuous economic variables poses a challenge in reinforcement learning because it requires handling an infinitely large state space. To address this, quantile-based discretization has been applied to categorize economic variables into bins, reducing complexity while maintaining essential trends.

Additionally, to distinguish between long-term economic trends and short-term fluctuations, the Hodrick-Prescott (HP) filter has been applied, which decomposes GDP into:

* Potential GDP (long-term trend component)
* Cyclical GDP (short-term fluctuations from the trend)

Using these components, we compute the output gap as follows:

Output Gap = x100

A large output gap suggests economic instability, prompting corrective monetary policy actions.

**Markov Decision Process (MDP) Transition Matrix**

The economic environment is modelled as a Markov Decision Process (MDP), where the system transitions between different economic states based on policy actions.

**State Space Representation**

Each state in the MDP is defined as:

*State = (Inflation Bin, Output Gap Bin, M2 Bin, Currency Ratio Bin)*

where:

* **Inflation Bin**: Captures inflationary pressure.
* **Output Gap Bin**: Measures economic performance relative to potential GDP.
* **M2 Bin**: Represents liquidity and potential inflationary impact.
* **Currency Ratio Bin**: Provides insight into cash circulation and informal economic activity.

By binning continuous variables into discrete categories, we create a manageable state space suitable for reinforcement learning.

**Action Space Representation**

The possible monetary policy actions include:

* Decrease Interest Rate (-1)
* Maintain Interest Rate (0)
* Increase Interest Rate (1)

Each action influences the state transitions, which are not entirely deterministic, making the system probabilistic in nature.

**MDP Transition Matrix Construction**

To estimate how states evolve, state transition probabilities has been calculated from historical data:

1. For each state-action pair, how often each next state occurs has been counted.
2. These counts have been normalized to obtain transition probabilities, forming the MDP transition matrix.
3. These probabilities guide Q-learning, making state transitions more realistic and structured.

**Q-Learning with MDP State Transitions**

Q-learning is a model-free reinforcement learning algorithm that helps the agent discover an optimal policy by interacting with the environment. The algorithm updates its knowledge using the Bellman equation:

*Q (s, a) = Q (s, a) + α (R + γ max Q (s′, a′) – Q (s, a))*

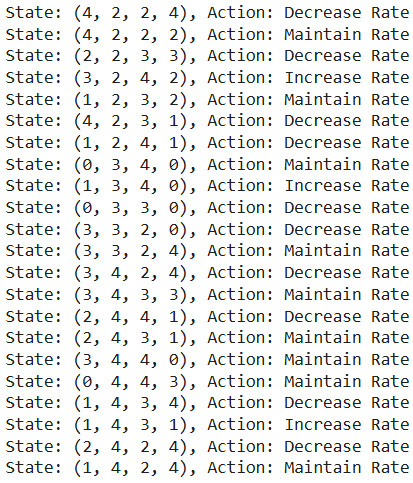
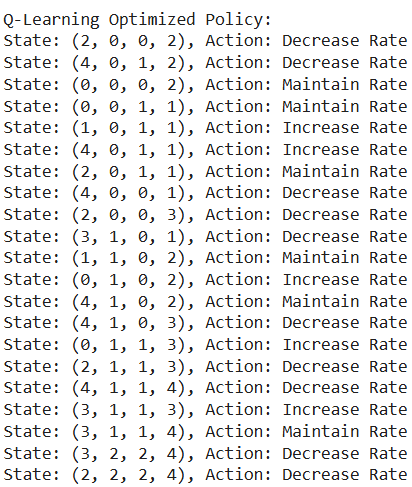
Where:

* α = learning rate
* γ = discount factor
* R = reward function
* s′ = next state
* a′ = next action

Incorporating MDP state transitions improves Q-learning by ensuring that next-state selections align with real-world transition probabilities, making the learning process more stable and efficient.

**Results**

The Q-learning optimized monetary policy demonstrates a data-driven strategy for interest rate adjustments based on the state of the economy. The learned policy suggests expansionary measures (decreasing interest rates) in recessionary conditions, contractionary measures (increasing rates) in inflationary environments, and neutral stances where the economy is stable. The results indicate that monetary policy actions must be adaptive, responding dynamically to fluctuations in inflation, output gap, liquidity, and cash circulation patterns.

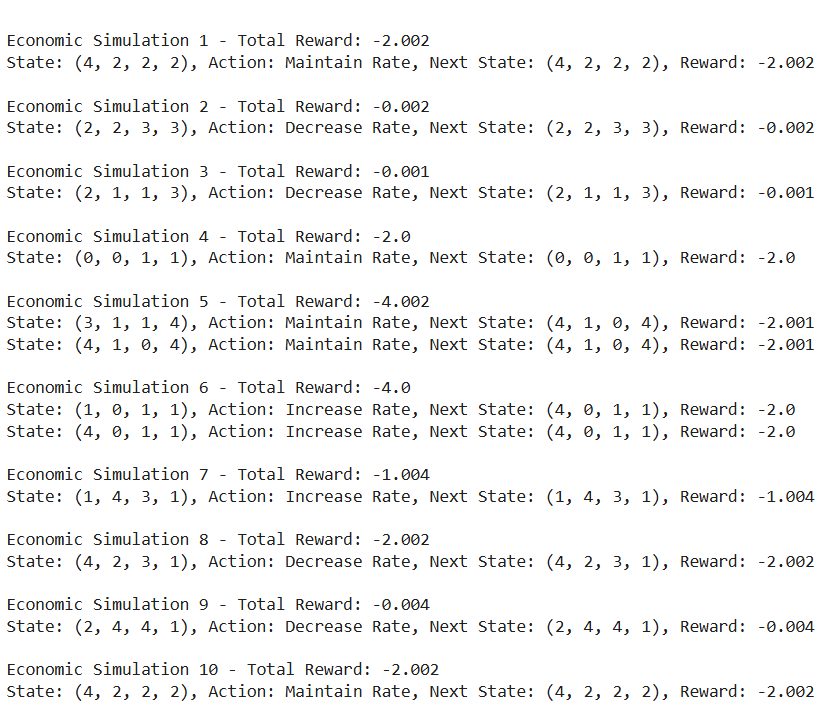


*Fig: Q-learning optimizes policies*

The State value defines the situation of the state variable:

* Inflation Bin (Index: 0): Represents inflation levels, where 0 = very low (deflation risk) and 4 = very high (inflationary pressure requiring rate hikes).
* Output Gap Bin (Index: 1): Measures economic performance; 0 = recession (below potential GDP) and 4 = overheating economy (above potential GDP, requiring tightening).
* M2 Bin – Money Supply (Index: 2): Captures liquidity; 0 = credit shortage, slow growth and 4 = excessive liquidity, risk of inflationary overheating.
* Currency Ratio Bin (Index: 3): Tracks cash circulation; 0 = highly formal economy and 4 = cash-dominant, high informal sector activity (weak policy transmission).

From the simulation results, several economic trends emerge. The model frequently suggests lowering interest rates when the output gap is negative and money supply is low, indicating an economy operating below its potential. Conversely, it recommends rate hikes in cases of excessive liquidity or high inflation, demonstrating alignment with traditional central banking strategies. The states that require policy intervention tend to cluster around extremes—either high inflation with liquidity excesses or low inflation with stagnation. These findings reinforce the importance of balancing economic growth and inflation control through monetary policy precision.



*Fig: Economic Simulations*

**Economic Interpretation of the Policies Formulated**

* **Expansionary Policy (Decreasing Interest Rates)**

One of the most significant findings is that the model frequently suggests lowering interest rates in states where the output gap is negative, inflation is moderate or low, and money supply is constrained. For instance, in state (2,0,0,2), the policy suggests a rate cut, indicating that the economy is experiencing low inflation, weak output, and restricted credit availability. This aligns with Keynesian economic principles, where lower interest rates boost demand, investment, and employment.

However, the model also carefully avoids aggressive rate cuts in cases where liquidity is already high. In state (3,1,0,1), for example, a rate decrease is recommended, but only when the output gap is sufficiently negative. This highlights the trade-off between stimulating growth and preventing excessive money supply, ensuring that inflation remains under control.

* **Contractionary Policy (Increasing Interest Rates)**

In states where inflation is high, liquidity is excessive, or the economy is overheating, the model suggests raising interest rates to prevent runaway inflation and asset bubbles. For instance, in state (4,0,1,1), where inflation is at its peak and the money supply is ample, the policy recommends a rate hike to slow down inflationary pressures. This follows the monetarist approach, where central banks use interest rate adjustments to control inflation by limiting credit expansion.

Another example is state (3,2,4,2), where inflation remains high, and the economy shows positive output growth. Here, the model correctly suggests an interest rate hike, preventing an overheated economy from spiralling into an inflationary bubble. This is aligned with the Taylor Rule, which suggests that interest rates should increase when inflation exceeds its target or when output exceeds potential GDP.

* **Neutral Policy (Maintaining Interest Rates)**

In several states, the model determines that no interest rate adjustments are necessary, implying that the economy is already in equilibrium. In states like (0,0,0,2) and (3,3,2,4), where inflation is low, the output gap is near zero, and liquidity conditions are balanced, the policy suggests maintaining interest rates.

This supports the idea that monetary policy should not always be interventionist. Frequent and aggressive rate changes can lead to market instability and policy credibility issues. The model’s ability to recognize when to take a hands-off approach demonstrates its robustness in balancing economic stability and policy intervention.

**Conclusion and Future Works**

The Q-learning optimized monetary policy provides a robust framework for interest rate decisions based on macroeconomic conditions. The results align well with real-world central banking principles, including:

* Lowering rates during recessions to stimulate demand.
* Raising rates in overheating economies to control inflation.
* Maintaining rates in stable conditions to prevent unnecessary intervention.

The model's ability to analyse inflation, output gaps, money supply, and currency circulation simultaneously highlights the complexity of monetary policy decisions. Unlike rule-based frameworks like the Taylor Rule, which use fixed formulas, this approach adapts dynamically to changing economic conditions.

Additionally, the model sheds light on the limitations of conventional monetary tools in high-cash economies, emphasizing the need for complementary fiscal policies and financial inclusion strategies.

Going forward, enhancing this model with real-time economic data and incorporating global macroeconomic shocks could improve policy responsiveness and accuracy. This AI-driven approach serves as a valuable decision-support system for central banks and policymakers, ensuring monetary policies remain data-driven, responsive, and stability-focused.

**References**

**Taylor, J. B. (1993).** *Discretion versus policy rules in practice.* *Carnegie-Rochester Conference Series on Public Policy, 39*, 195-214.

**Bernanke, B. S., & Blinder, A. S. (1992).** *The Federal Funds Rate and the Channels of Monetary Transmission.* *The American Economic Review, 82*(4), 901-921.

**Clarida, R., Galí, J., & Gertler, M. (1999).** *The Science of Monetary Policy: A New Keynesian Perspective.* *Journal of Economic Literature, 37*(4), 1661-1707.

**Wang, Z., & Yang, J. (2020).** *Application of Reinforcement Learning in Financial Decision Making.* *Finance Research Letters, 36*, 101569.