AMRO
ASEAN +3 MACROECONOMIC RESEARCH OFFICE

Working Paper (WP/25-13)

Can an AI Agent Hit a Moving Target?

Aruhan Rui Shi

November 2025

[This page is intentionally left blank]

Can an AI Agent Hit a Moving Target?

Aruhan Rui Shi*
Authorized by Abdurohman (Deputy Director)

November 2025

Abstract

Policymakers face challenges in understanding economic agent behavior during structural transformations, where traditional rational expectations (RE) models often struggle to capture the complexities of adaptive decision-making. This research explores how integrating artificial intelligence (AI) into a general equilibrium framework can enhance theoretical modeling by representing adaptive behaviors that integrate perspectives from neuroscience and psychology. Using a deep reinforcement learning (RL) approach, the model highlights how agents explore options, balance objectives, and adjust strategies during an economic regime change, such as an acceleration in the money supply process. Simulation results illustrate that AI agents, guided by exploration-driven learning, adapt their consumption, savings, and liquidity holding decisions in response to structural changes. These agents show a degree of convergence toward RE outcomes while retaining the flexibility to adjust under dynamic conditions, capturing behaviors that traditional models may overlook. This study provides a structured framework for analyzing bounded rationality in economies undergoing structural changes. It offers a complementary perspective to conventional approaches and highlights new avenues for research on adaptive policy design.

JEL Codes: C45, D83, D84, E31, E50

Keywords: expectations formation, artificial intelligence, reinforcement learning, bounded rationality, monetary policy, structural change

*I am grateful for the valuable advice and support of Roger Farmer and Herakles Polemarchakis. I am thankful for the insightful comments from Martin Ellison and Pablo Beker. I appreciate the financial support from Warwick University and the many conversations with colleagues at the Rebuilding Macroeconomics project. I am grateful for the helpful discussions and guidance from colleagues at the Bank of England, which also inspired the choice of the DDPG algorithm I used in this exercise. I thank participants at the 2021 Money Macro Finance annual conference, the Society for Computational Economics 28th International Conference, the CES 2022 annual conference, the 2022 Barcelona School of Economics PhD Workshop on Expectations in Macroeconomics, and the 2nd ASEAN+3 Finance Think-tank Network (AFTN) Seminar at the Hong Kong Monetary Authority. All remaining errors are mine. Please address correspondence to Shi, aruhan.rui.shi@amro-asia.org, aruhanruishi@outlook.com (permanent).

Contents

List of Figures

List of Tables

[This page is intentionally left blank]

1 Introduction

The modeling of adaptive economic agents during periods of structural changes has been a topic of extensive discussion, dating back to the accelerationist debate (Sargent, 1971). The accelerationist argument, characterized by a backward-looking Phillips curve combined with adaptive expectations, suggests a plausible trade-off between inflation and unemployment. This theoretical construct posits that a government could sustain low unemployment by accelerating the money supply process. It raises the fundamental question of whether policymakers can exploit agents with naive adaptive expectations who consistently make errors.

In response to this debate, the rational expectations (RE) hypothesis was introduced, assuming that agents form model-consistent beliefs and possess a deep understanding of the economy. The RE hypothesis builds on the works of Lucas (1972, 1976); Sargent (1971); Sargent and Wallace (1973). The shift from naive to more sophisticated agents sparked a revolution in economic modeling, offering considerable advantages, particularly for conducting policy experiments in stable environments, where the underlying data-generating processes or model structures remain unchanged. However, RE agents lack adaptability in dynamic environments, where structural changes occur. Therefore, the exploration of modeling adaptive agents and understanding their behaviors during such changes remains an important and contentious issue.

Economic agents, or in reality, human beings, exhibit a level of intelligence that falls between that of RE agents and naive adaptive expectations agents. Their expectations and behaviors are intricately nuanced. They engage in exploration of different options, evaluate their actions, draw connections between the present and the past, consider long-term satisfaction, and may also prioritize immediate gratification. When faced with changes in their environment, they demonstrate the ability to adapt (Bassett and Mattar, 2017).

Artificial intelligence (AI) technologies, particularly deep reinforcement learning (RL), have proven effective in enabling artificial agents to learn specific tasks. Core ideas in RL are inspired by phenomena in animal learning, psychology and neuroscience (Subramanian et al., 2022).[1] The RL framework dates back to the 1950s and 1960s, when it combined trial-and-error learning with the formalized Bellman equation. In recent years, it has gained traction through successful applications in fields such as chemistry, robotics, and natural language processing. Notable examples include the development of large language models like ChatGPT and ClaudeAI, as well as breakthroughs in protein structure prediction, for which the researchers behind the AI application were awarded the 2024 Nobel Prize in Chemistry.

Given the limited use of AI technologies as a framework for modeling adaptive economic agents

[1] For example, Schultz et al. (1997) find that the core of temporal-difference (TD) method can be directly related to actual biological phenomena. Specifically, changes in dopamine levels in the brain—transmitted from certain regions to various target areas—correspond to TD reward prediction errors. The TD method will be introduced in Section 4.

and the complexity inherent in AI frameworks, my objective is to evaluate the ability of an AI agent to adapt and modify its behavior in response to significant environmental changes within a simple dynamic stochastic general equilibrium model. Applying the AI framework aims to provide a complement to the naive form of adaptive expectations and RE while addressing the constraint of the RE hypothesis in adapting to structural changes. This change is marked by a shift in the nominal money supply, resulting in the aggregate price sequence transitioning from a stationary state to a non-stationary one. I focus on how well the AI agent adjusts its beliefs and decisions and examine the impact of varying levels of exploration on its adaptive behavior during such an economic shift. These investigations are crucial for demonstrating the practical applicability of AI algorithms in modeling agents with bounded rationality in economies undergoing structural transformations.

I propose to model the adaptive behaviors of an economic agent during a structural change using AI technology, specifically a deep RL framework, for two reasons:

1. AI agents are designed to adapt to structural changes, such as an acceleration in the money supply process, due to their exploration capabilities. This contrasts with RE agents, who excel at a single task and operate based on fixed beliefs, whereas AI agents are designed to be adaptive across various tasks with evolving beliefs.[2]

2. AI agents learn from experience, providing a theoretical framework for modeling experience-based learning.[3] Notably, Malmendier and Nagel (2016) highlight the role of experience in shaping inflation expectations, effectively capturing behaviors observed in microdata.

To conduct the experiment on regime change, I employ a simple dynamic stochastic general equilibrium model that incorporates transaction costs of money demand and utilize a representative AI agent operating according to the deep RL framework. Drawing inspiration from the accelerationist debate, I design the monetary authority to increase the daily growth rate of the nominal money supply from 0% to 1%. The representative AI agent makes two key decisions in each period: the storage decision, determining the quantity of real resources to store (and consume), and the liquidity holding decision, which influences the desired level of real balances to minimize transaction costs. Within the deep RL framework, the agent's policy and value functions are approximated using artificial neural networks (ANNs), allowing flexible adjustment of their forms and parameters.

Through simulation experiments, I find that:

1. The AI agent's capability to adapt its beliefs in response to economic changes is evident in its adjustments to consumption, storage, and demand for real balances. The level of exploration

[2] Along with an extensive review of the literature on the applications of the deep RL framework in economics, Shi (2023) show that this framework involves training AI agents to interact with the economy and take exploratory actions, enabling them to learn from these interactions and adjust their policy and value functions over time.

[3] Hertwig and Erev (2009) discuss the significance of experience in the decision-making process from the perspective of cognitive sciences.

significantly affects this adaptability, which in turn impacts aggregate economic transitions, as shown by the inflation levels following an increase in the nominal money supply. Specifically, after the structural change, inflation rises to 1% in the high-exploration economy, while in the low-exploration economy, it stabilizes around 1.5%.

2. As aggregate inflation increases, the total real resources available for AI agents to allocate across decisions—consumption, storage, and liquidity holdings—decrease. Since both agents maintain similar consumption levels in both regimes, they need to reduce their storage and liquidity holdings in the new regime, where fewer real resources are available. The high-exploration agent reduces its liquidity holdings by 25%, more than the low-exploration agent's 22%. As a result, to maintain its consumption level, the low-exploration agent must also reduce its storage, leading to a decrease in wealth. In the new regime, the low-exploration agent accumulates 0.6% less wealth than the high-exploration agent.

3. Due to their exploration feature, the AI agents' beliefs adjust toward those of the RE agent in both regimes but do not fully align. For example, in the new regime, their liquidity holdings decrease compared to the initial regime, similar to the RE agent, but not by the same amount. However, unlike the RE agent, they demonstrate the ability to adjust and adapt to a regime change.

This research provides theoretical insights for policymakers and offering an initial step toward modeling complex adaptive behaviors as a complement to traditional RE models. The findings highlight the importance of incorporating agents' adaptive behaviors and exploration-driven learning into macroeconomic models for economies undergoing structural changes. Different adaptive behaviors can lead to varying transitional dynamics and welfare disparities during such changes.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the economic environment. Section 4 introduces the AI framework and explains its integration with the economic model. Section 5 outlines the parameters used. Section 6 presents the experiments and results. Finally, the paper concludes in the last section.

2 Related Literature

This paper contributes to the literature on modeling agents' expectation formation processes, beginning with the concept of adaptive expectations. Keynes emphasized the importance of agents' expectations when he demonstrated how they influence output and employment (Keynes, 1936). Two decades later, Cagan (1956) and Friedman (1957) formalized the idea of adaptive expectations. Combined with the Phillips curve, their work sparked a significant debate about whether and how a government could exploit the potential negative relationship between inflation and unemployment.

However, their approach was criticized for assuming that agents would forecast inflation solely based on the previous period's inflation.

Proposed as an alternative to adaptive expectations was the rational expectations hypothesis, which assumes that agents form model-consistent beliefs and have an understanding of the economy (Lucas, 1976; Sargent, 1971; Sargent and Wallace, 1973). One of its key advantages is its usefulness in analyzing policy experiments in a stationary environment. However, it also has drawbacks, including its inability to provide convincing inflation dynamics in response to shocks. Various techniques have been proposed to strike a middle ground between rational expectations, where agents are too smart, and adaptive expectations, where agents are too naive.

Due to the limitations of the RE framework, an alternative approach proposed was information rigidities, which includes sticky information (Mankiw and Reis, 2002; Ball et al., 2005), noisy information (Woodford, 2001) and rational inattention (Sims, 2003). The main idea is that agents are constrained in obtaining or processing information, and therefore use only a portion of the available information to make 'optimal' decisions. In other words, they still hold model-consistent beliefs but with less than full information.

Another approach proposed focuses on bounded rationality (Sargent, 1993) and adaptive learning (Evans and Honkapohja, 1999). Schorfheide (2005), Ozden and Wouters (2021), and Airaudo and Hajdini (2021) examine the combination of adaptive learning with Markov switching specifications to model learning agents in the context of policy regime changes. In the adaptive learning literature, agents are considered as knowledgeable as econometricians, learning about model parameters by running regressions on past data or using Bayesian updating.

AI agents in this paper share both similarities and differences with the two alternatives proposed as complements to the RE hypothesis. I argue that agents are limited in the amount of information they can collect and process at any given time. However, unlike traditional models, the agents in my approach do not form model-consistent beliefs; instead, they develop decision-making strategies based on their own experiences. Modeling decision-making processes using AI algorithms is also a form of adaptive learning, as it relies on learning from past experience. However, AI agents differ in that they learn from their own interactions with the environment by making exploratory actions.

This paper is similar to recent literature that uses ANNs to model economies. Ashwin et al. (2021) study the stability properties of multiple equilibria with learning agents, where the agents learn using ANNs. Similarly, Kuriksha (2021) models economic agents with deep ANNs in a macro-financial environment. My approach differs in that my learning agents generate their own experiences by interacting with the environment, whereas both Ashwin et al. (2021) and Kuriksha (2021) primarily use deep learning methods (deep ANNs).

This paper is closely related to the emerging literature on the application of deep RL algorithms in

macroeconomic models. In Shi (2021), they use a stochastic optimal growth model, and show that an AI agent can learn without prior information about the economic structure or its own preferences, and can adapt to both transitory and permanent income shocks. Recent studies increasingly explore deep RL algorithms as solution methods. Chen et al. (2021) implement a deep RL algorithm in a model with different monetary and fiscal policy regimes, showing that a deep RL agent can locally converge to all equilibria. Hinterlang and Tänzer (2021) employ a deep RL algorithm to solve for optimal policy responses. Additionally, Hill et al. (2021) and Curry et al. (2022) investigate the use of deep RL algorithms in solving multi-agent macroeconomic models.

3 The Model: an Economist's Approach

I employ a transaction cost of money demand model to illustrate the role of money in the economy, akin to Sims (1994). In this model, the preference of individuals for holding money stems from its capacity to diminish transaction costs. Although there are alternative methods to incorporate money into economic models, such as direct integration into utility functions, I opt for the transaction cost approach due to its intuitive appeal. It is important to recognise that money itself does not inherently provide utility; rather, its value derives from its ability to facilitate the purchase of goods and services and its convenience in enabling transactions. In this model, an AI agent is required to determine a policy within a framework that encompasses decisions on both storage and real balance in each period. This represents a departure from Shi (2021), where the agent's decisions were confined to savings each period.

By integrating money into the model, the monetary authority acquires policy leverage to influence the decisions of private agents through changing the growth rate of the nominal money supply. Consequently, this model offers a platform to conduct experiments to explore the behavioral shifts of AI agents in response to a regime change in monetary policy, contrasting with Shi (2021), which focused solely on the real sector of the economy.

I first illustrate how the model is conventionally presented in economics as a benchmark.

3.1 A Representative Household

A representative household aims to maximise its lifetime utility, as outlined in Equation 1.

$$E_0 \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{1}$$

subject to the constraint,

$$c_t(1 + \frac{1}{\kappa} f(v_t)) + s_{t+1} + \frac{M_{t+1} - M_t}{P_t} \leq y_t + s_t^{\alpha} + \tau_t \tag{2}$$

where $\beta \in (0,1)$, P_t is the price level at period t, c_t denotes the consumption at t, M_t is nominal money balance, s_t is the saved stock of goods that a household enters period t with, and there is a storage technology that takes the Cobb-Douglas form. y_t is the endowment or income of the agent. τ_t is the government transfer at t.

$\frac{1}{\kappa}f(v)$ represents transactions costs per unit of consumption. $f(v_t) = v_t$, and velocity is $v_t \equiv \frac{c_t P_t}{M_{t+1}}$. Real balance is defined as $m_{t+1} = \frac{M_{t+1}}{P_t}$. Demanding or holding real balance reduces transaction cost.

The endowment depends on a constant $\bar{y}$ and an exogenous process ϵ_t^y, where ϵ_t^y is drawn randomly from a normal distribution with zero mean and variance of 0.01.

$$y_t = \bar{y} + \epsilon_t^y \tag{3}$$

3.2 Government: Fiscal and Monetary Policies

The government conducts an active monetary policy and a passive fiscal policy. The monetary authority follows a money growth rule:

$$M_{t+1} = \delta^M M_t, \tag{4}$$

where δ^M is a policy variable that determines the speed of money supply. In real terms, this money supply rule becomes,

$$m_{t+1} = \delta^M \frac{m_t}{\pi_t}, \tag{5}$$

where real balance $m_t \equiv \frac{M_t}{P_{t-1}}$.

Government budget constraint is,

$$g + \tau_t = \frac{M_{t+1} - M_t}{P_t}. \tag{6}$$

The government sets initial nominal money supply and a policy variable δ^M. It also determines a constant government spending g.

Combine the household and the government budget constraints equation 2 and 6, the consolidated budget constraint is

$$c_t(1+\frac{1}{\kappa}v_t)+s_{t+1}=y_t+s_t^{\alpha}-g. \tag{7}$$

3.3 Household's First Order Conditions (FOCs)

The household's Lagrangian is specified as

$$\mathcal{L}=E_t\sum_{t=0}^{\infty}\beta^t\left\{ln(c_t)+\lambda_t\left[y_t+s_t^{\alpha}+\tau_t-c_t(1+\frac{1}{\kappa}v_t)-s_{t+1}-\frac{M_{t+1}-M_t}{P_t}\right]\right\} \tag{8}$$

where $v_t=\frac{c_tP_t}{M_{t+1}}$.

The FOC with respect to consumption c_t is ,

$$\frac{1}{c_t}=\lambda_t(1+\frac{2v_t}{\kappa}). \tag{9}$$

The FOC with respect to M_{t+1} is,

$$\frac{\lambda_t}{P_t}(1-\frac{v_t^2}{\kappa})=\beta E_t\frac{\lambda_{t+1}}{P_{t+1}}. \tag{10}$$

The FOC with respect to s_{t+1} is,

$$\lambda_t=\alpha\beta s_{t+1}^{\alpha-1}E_t\lambda_{t+1}. \tag{11}$$

3.4 Non-Stochastic Steady States and Determinancy

The equations governing the dynamic system of this model consist of the first-order conditions, specifically Equations 9, 10, and 11, alongside the consolidated budget constraint, the rule governing the money supply, and the process determining the endowment. Additionally, I incorporate the definition of velocity directly into these equations and use inflation instead of price levels. Inflation is defined as $\pi_t=\frac{P_t}{P_{t-1}}$.

$$\left[\frac{1/c_t}{1+\frac{2c_t}{\kappa m_{t+1}}}\right]\left[1-\frac{c_t^2}{\kappa m_{t+1}^2}\right] = \beta E_t \frac{\frac{1/c_{t+1}}{1+\frac{2c_{t+1}}{\kappa m_{t+2}}}}{\pi_{t+1}} \quad (12)$$

$$\left[\frac{1/c_t}{1+\frac{2c_t}{\kappa m_{t+1}}}\right] = \alpha\beta s_{t+1}^{\alpha-1} E_t \left[\frac{1/c_{t+1}}{1+\frac{2c_{t+1}}{\kappa m_{t+2}}}\right] \quad (13)$$

$$m_{t+1} = \delta^M \frac{m_t}{\pi_t} \quad (14)$$

$$c_t(1+\frac{c_t}{\kappa m_{t+1}}) + s_{t+1} = y_t + s_t^\alpha - g \quad (15)$$

$$y_t = \bar{y} + \epsilon_t^y \quad (16)$$

To solve for the non-stochastic steady state, I assume zero shock and constant real variables. The non-stochastic steady state values are presented in Table 5, given the specified parameters.

I arrange the equations in the form of Equation 17.

$$E_t f\left\{Y_{t+1}, Y_t, X_{t+1}, X_t\right\} = 0 \quad (17)$$

where E_t denotes the expectations operator conditional on information available at time t. The state vector is $X_t = [s_t, y_t, m_t]'$. The co-state vector is $Y_t = [c_t, \pi_t]'$.

The steady state is locally determinate and unique. I check whether the non-stochastic steady state is locally determinant by deriving the Jacobians of this system, as written in the form of Equation 17. I calculate generalized eigenvalues using QZ decomposition, all in accordance with the parameters stated in Section 5. I calculate the list of eigenvalues to be $[0, 1, 0.18, 5.48, 0]$. There are three stable roots, two unstable ones, and two co-states are solved from three state variables. Therefore, I conclude that the steady state is unique and locally determinate. The same exercise is repeated for the second regime with an increasing supply of nominal balance, leading to the same conclusion: the steady state is locally determinate and unique.

4 The Model: an AI Approach

I begin this section by introducing the AI learning framework, outlining its key components and highlighting the importance of the exploration feature. This is followed by the implementation details for the specific model presented in Section 3, along with a full description of the algorithm.

4.1 An AI Learning Framework: a Brief Introduction

RL is a type of AI technology where agents learn by interacting with an environment, taking actions, and receiving rewards based on their actions. Deep RL extends this concept by using deep neural networks to approximate value or policy functions, which guide the agent's decision-making process. In RL, the learning process relies on the agent generating data by interacting with the environment, using these experiences to improve its performance and solve complex problems. Sutton and Barto (1981) provide a comprehensive review of RL algorithms.

Unlike supervised learning, which requires labeled training data to evaluate its predictions, or unsupervised learning, which seeks to find patterns in unlabeled data, RL does not need a predefined training dataset. Instead, it learns through continuous interaction with its environment. The agent receives feedback via a reward function, which informs it whether its actions are beneficial or not, helping it optimize its behavior over time.

RL originated from early studies in psychology and neuroscience, where it was initially developed to explain how animals learn through experience. As detailed in Tohid and Shi (2022), pioneers such as Alan Turing, Norbert Wiener, and Richard Bellman initially explored these ideas, which later gained significant traction in computer science during the 1970s. Two key concepts underpinning RL's development were temporal difference (TD) learning, which involves trial-and-error learning, and optimal control, rooted in dynamic programming. RL reached the machine learning community in the late 1980s, leading to breakthroughs such as Watkins' Q-learning algorithm. By the 1990s, RL was applied to real-world problems like backgammon, achieving human-level performance, and began to merge with techniques like neural networks and evolutionary computation. By the late 2000s, the integration of RL with deep learning facilitated applications in complex fields such as robotics, healthcare, and finance, significantly advancing the field.[4]

4.2 An AI Learning Framework: the Actor-Critic Model

The specific deep RL algorithm adopted here was first introduced by Lillicrap et al. (2015), namely deep deterministic policy gradient (DDPG). [5] This algorithm mainly follows the actor-critic model of reinforcement learning, and it uses the formal framework of a Markov Decision Process (MDP) to define the interactions between a learning agent and its environment in terms of states, actions, and rewards (Figure 1).

State S is a random variable from a state space, which is a bounded and compact set,[6] i.e., $S \in \mathcal{S}$.

[4]For example, machines trained with the deep Q-network algorithm (Mnih et al., 2013) achieved human-level performance on many Atari video games using raw pixels as input. However, the deep Q-network algorithm is limited to discrete action spaces. For continuous action spaces, the deep deterministic policy gradient (DDPG) algorithm from Lillicrap et al. (2015) is often employed.

[5]Chen et al. (2021) also adopt the DDPG algorithm in their study of learnability of rational expectations equilibrium in different policy regimes.

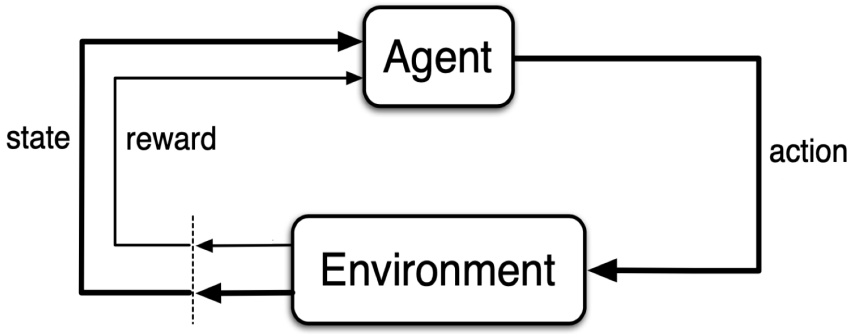
[6]The latest research on reinforcement learning also investigates the setting with unbounded state space, e.g., Shah

The agent takes an action A, which belongs to an action space $\mathcal{A}$, $A \in \mathcal{A}$. The state evolves through time following a probability function, $p : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$, which is defined as

$$p(S'|S, A) \equiv Pr\{S_t = S'|S_{t-1} = S, A_{t-1} = A\}. \tag{18}$$

It shows the probability of a random variable state S' occurring at time t, given the preceding values of state, S, and action, A.

Figure 1: The agent-environment interaction in a reinforcement learning setting



Source: Sutton and Barto (2018)

Reward is a random variable and can be generated from a reward function, $R : \mathcal{S} \times \mathcal{A} \to \mathcal{R}$.

The return from a state is defined as the sum of discounted future rewards,

$$G_t \equiv R_t + R_{t+1} + ... = \sum_{k=0}^{\infty} \beta^k R_{t+k}, \tag{19}$$

where β is the discount factor.

A RL learning agent's behaviors follow a policy function, also known as the actor network. The policy function can be both stochastic and deterministic. A stochastic policy maps states to probabilities of selecting each possible action. A deterministic policy, which is used in this paper, maps a state from a state space to an action from an action space, and it is denoted by $\mu : \mathcal{S} \to \mathcal{A}$.

A value function[7], also known as the critic network, shows the 'expected' returns of taking an action in a state and thereafter following policy μ. Expectations here are subjective beliefs that depend on

et al. (2020).

[7] In reinforcement learning literature, two types of value functions are defined: action-value function and value function. To not complicate the matter, I use value function as action-value function throughout this paper.

learning agents' past experiences. A value function is defined as,

$$Q^{\mu}(S, A) \equiv E^{\mu}[G_t | S_t = S, A_t = A], \tag{20}$$

where Q^{μ} means the action value function follows policy μ, and E^{μ} reflects that it's a subjective belief that depends on a policy μ that is formed by past experience.[8]

Many approaches in reinforcement learning make use of the recursive relationship known as the Bellman equation,

$$Q^{\mu}(S, A) = R(S, A) + \beta E^{\mu} Q(S', A'), \tag{21}$$

where $A' = \mu(S')$.

RL methods focus on how the learning agents' policy and value functions change as a result of their experience. These changes can be a functional form change or parameter value updates. The DDPG algorithm uses ANNs to approximate policy and value functions: the actor network is denoted as $\mu(S|\theta^{\mu})$, where θ^{μ} represents parameters of the ANNs; the critic network is denoted as $Q(S, A|\theta^{Q})$, and θ^{Q} is its parameters. θ^{μ} and θ^{Q} are updated during learning, and can be viewed as the coefficients of two functions and the probabilities involved in making subjective expectations. Two ANNs are updated with respect to each other.

The goal of RL learning agents is to continuously update their subjective beliefs about the world based on experience, and to form a decision-making strategy (approximated by the actor network) that produces the highest discounted future returns (approximated by the critic network). The actor network is updated with the goal of maximizing the corresponding critic network. In other words, the actor network is updated based on what the agent believes, at that time, to be a strategy that produces high 'expected' returns. The critic networks evolves over time. What the agent follows as the critic network at period t is different from what it is at $t+1$. Expectations are the learning agents' subjective beliefs that are formed from past experience.

The critic network is updated with the goal of minimising a TD error (a temporal difference error).[9] The TD error follows the form,

$$T - Q(S, A|\theta^{Q}), \tag{22}$$

where T is called a TD target, and it adds the reward given a state-action pair to the discounted values of the next state and action, i.e.,

[8] This forms of notation, e.g., E^{μ}, largely follows The Handbook for Reinforcement Learning by Sutton and Barto (2018).
[9] The full algorithm is in the next section.

$$T = R(S, A) + \beta Q(S', A'|\theta^Q) \tag{23}$$

and the next period action A' is assumed to follow the actor network $\mu(S|\theta^\mu)$ at that time.[10]

$$A' = \mu(S'|\theta^\mu). \tag{24}$$

Intuitively, TD targets represent the best possible returns learning agents receive following a state-action pair and their subjective beliefs.

Neural science research shows that the dopamine neuron firing rates in the brain resemble the TD error sequence during learning (Botvinick et al., 2019). This motivates research in neural science to model decision-making in connection with RL algorithms.

4.3 An AI Learning Framework: Exploration

Explorations play a crucial role because learning agents that make exploratory actions collect a wide range of information. An exploratory policy is defined as,

$$\mu'(S_t) = \mu(S_t|\theta^\mu) + \mathcal{N}_t. \tag{25}$$

This shows that the final action the agent takes, i.e., what $\mu'(S_t)$ generates, depends on the actor network $\mu(S_t|\theta^\mu)$, and a random variable sampled from a noise process $\mathcal{N}_t$. Following Lillicrap et al. (2015), $\mathcal{N}_t$ is sampled from a discretised Ornstein-Uhlenbeck (OU) process.[11]

This exploratory policy produces random actions. The randomness decreases over time (by design) but it never disappears. The implication is that in a stationary environment, the agents learn their policy functions, and the learnt functions can be similar to the true policy function but may never be identical. However, in a non-stationary environment, explorations allow the policy network to adjust and be flexible to changes in the environment.

The exploration strategy implies that the policy function can converge to a close region of the rational expectation solution (if it exists), but will not be identical to it. In an environment with structural breaks or regime changes, this exploratory policy allows the learning agent to adjust its expectations and adapt its policy to a new regime. While my focus is on the exploration-driven adaptive learning framework provided by the DDPG algorithm, many RL algorithms aim to address dynamic programming problems with reduced computational effort and without relying on a perfect model of the environment (Sutton and Barto, 2018).

[10] It need not be the same as the true policy.

[11] There is a strain of literature in computer science that solely focuses on different exploration strategies to achieve the best performance for a given task. It is out of the scope of the current exercise, and not discussed here.

4.4 Bridge AI and Economics

To apply and implement the AI framework, the economic model is first translated into components of a MDP, including state, action, reward, and next state[12]. These components are summarized in Table 1. The state variables encompass current period income and storage technology. The agent is also aware of past period inflation, real balance, and storage level.

The actions available to the AI agent consist of determining the proportion of real resources to store, λ_t^s, and the desired level of real balance to demand, m_{t+1}. Based on its storage decision, the agent allocates resources for consumption. Likewise, its decision on real balance demand affects transaction costs and aggregate inflation dynamics. In this representative agent model, the AI agent's demand for real balance represents the aggregate money demand, although in multi-AI agent studies, aggregate money demand would differ from individual demands for real balance.

[12]To avoid confusion with the storage variable s_t, I use capital letters to denote deep RL-related terms: state (S_t), action (A_t), and reward (R_t).

Table 1: RL Components and the Economic Environment

Terminologies	**Description**	**Representation in the economic environment**
State, S_t	A random variable from a state space, $S_t \in \mathcal{S}$	$S_t = \{y_t, \pi_{t-1}, m_t, s_t\}$
Actions, A_t	A random variable from an action space, $A_t \in \mathcal{A}$	$A_t = \{\lambda_t^s, m_{t+1}\}$
Rewards, R_t	A function of state and action	$R_t = ln(c_t)$
Policy function, $\mu(S\|\theta^\mu)$	A mapping from state to action, $\mu : \mathcal{S} \rightarrow \mathcal{A}$	Approximated by a neural network, i.e., actor network; parameterised by θ^μ to be updated during learning
Value function, $Q(S, A\|\theta^Q)$	The 'expected' (subjective belief) return from taking an action in a state	Approximated by a neural network, i.e., critic network; parameterised by θ^Q to be updated during learning

The reward function, $R_t = ln(c_t)$, is influenced by the agent's level of consumption, c_t. Higher consumption leads to greater rewards per period. Policy ($\mu(S|\theta^\mu)$) and value ($Q(S, A|\theta^Q)$) functions are approximated using two deep neural networks. The AI agent aims to develop a decision-making strategy that maximizes its performance not only in each period but also over the long run, as measured by cumulative rewards. The agent seeks consistent high-level rewards rather than focusing solely on maximizing rewards in a single period.

Consequently, the agent faces two critical decisions: 1) determining the balance between storage and consumption, and 2) managing transaction costs through the allocation of resources to hold money. Higher storage allows for increased future consumption when adverse income or endowment shocks occur, though it means consuming less in the current period. The second decision involves optimizing the allocation between money and real goods. Holding more money reduces transaction

costs but limits the real goods available for consumption or storage. Conversely, holding less money frees up more real goods for consumption or storage but increases transaction costs.

4.5 Full Algorithm and Sequence of Events

The complete algorithm comprises three main steps: initialization, interaction, and learning.

Step I. Initialization

- In a given environment, design a state space $\mathcal{S}$, a continuous bounded and compact set for random variables specified in Table 1; design an action space $\mathcal{A}$, a continuous bounded and compact set for the action (random) variables.
- Set up two deep ANNs: an actor network $\mu(S|\theta^{\mu})$ takes the argument of a state from the state space and generates an action within the action space; a critic network $Q(S, A|\theta^{Q})$ takes the argument of a realised state-action pair and generates a value. Setting up two ANNs involves determining the input and output dimensions, the specific architectures, the number of layers, the number of nodes per layer, and how nodes are connected.
- θ^{μ} represents the parameters of the actor network, and θ^{Q} represents the parameters of the critic network.
- Define a memory $\mathcal{B}$ (called replay buffer or transitions in the deep RL literature), which stores information that is collected by a deep AI agent during the agent-environment interactive process. One period memory (or a transition) is characterised by a sequence of variables (S_t, A_t, R_t, S_{t+1}).
- Define a length of N, which is the size of a mini-batch. A mini-batch refers to a sample drawn from the memory, $\mathcal{B}$.
- Define the total number of episodes E and simulation periods per episode. The higher the episodes, the longer the learning periods.[13]
- Lastly, a simulation period of T is established for each episode, with T exceeding the value of N.

For each episode, define the initial state,[14] and loop Steps II and III.

[13] In the deep RL literature, the AI agent is usually set to learn a particular task or an Atari game. An episode thus means re-starting the game or task, and it ends with a terminal state (i.e., the end result of a game). In an economic environment, however, a clear terminal state can be difficult to specify. Therefore, the concept of episodes only correlates to how long an agent has been learning.

[14] Initial state variables can also be randomly drawn from the state space.

Step II. Interaction: the AI agent starts to interact with its environment for $t \leq T$. This step involves how agents' actions are chosen and how their actions impact the aggregate economy.

- Assume $\delta^M = 1.00$ (δ^M=1.01 in the new regime) and other variables that are known to the agent at period t: y_t, π_{t-1}, m_t, s_t.
- The agent selects a random (based on the randomly initialised actor network) of action variables, $A_t = \mu(S_t|\theta^\mu) + \mathcal{N}_t$, and A_t contains λ_t^s and m_{t+1}, and $\mathcal{N}_t$ is a noise attached to the action to ensure exploration, which is sampled from an AR(1) process.
- Given the exogenous (to the AI agent) nominal money supply from the government, $M_{t+1} = \delta^M M_t$, with aggregate money demand equal to aggregate money supply, price level or inflation can be derived as $\pi_t = \frac{M_{t+1}}{M_t}\frac{m_t}{m_{t+1}} = \delta^M \frac{m_t}{m_{t+1}}$. In a one-agent case, aggregate money demand at period t is m_{t+1}.
- The amount stored is $s_{t+1} = \lambda_t^s(y_t + \frac{m_t}{\pi_t} + s_t^\alpha + \tau_t - m_{t+1})$.
- c_t is reached from the budget constraint.
- The new state variables are $S_{t+1} = \{y_{t+1}, \pi_t, m_{t+1}, s_{t+1}\}$, where $y_{t+1} = \bar{y} + \epsilon_{t+1}^y$, and ϵ_{t+1}^y is sampled from a normal distribution $N(0, 0.01)$, $\bar{y} = 1$.
- The reward the agent receives is, $R_t = u(c_t)$.
- Store a transition (S_t, A_t, R_t, S_{t+1}) in the memory $\mathcal{B}$.

Step III. Learning: training the AI agent (when the AI agent starts to learn) for period $N \leq t \leq T$.

- Sample a random mini-batch of N transitions (S_i, A_i, R_i, S_{i+1}) from the memory $\mathcal{B}$.
- Calculate the TD-target values TD_i for each transition $i \in N$ following

$$TD_i = R_i + \beta Q^\mu(S_{i+1}, \mu(S_{i+1}|\theta^\mu)|\theta^Q) \tag{26}$$

  where $Q^\mu(S_{i+1}, \mu(S_{i+1}|\theta^\mu)|\theta^Q)$ is a prediction made by the critic network with state-action pair $(S_{i+1}, \mu(S_{i+1}|\theta^\mu))$, and $\mu(S_{i+1}|\theta^\mu)$ is a prediction made by the actor network with input S_{i+1}.
- Obtain $Q(S_i, A_i|\theta^Q)$ from the critic network with input state-action pair (S_i, A_i)
- Calculate the average loss for this sample of N transitions

$$L = \frac{1}{N}\sum_i \left(TD_i - Q(S_i, A_i|\theta^Q)\right)^2 \tag{27}$$

- Update the critic network with the objective of minimising the loss function L.[15]
- For the policy function, i.e., the actor network, the objective is to maximise the value function predictions. Define the objective function as,

$$J(\theta^{\mu}) = Q^{\mu}(S_i, \mu(S_i|\theta^{\mu})|\theta^{Q}). \tag{28}$$

- Maximising this objective function is equivalent to minimising $-J(\theta^{\mu})$. Update the actor network parameters θ^{μ} with the objective of minimising $-J(\theta^{\mu})$.[16]

5 Parameters and (Non-Stochastic) Steady State Values

I choose the discount factor to be 0.99999 so that each simulation period corresponds to a day. The rest of the parameters are uncalibrated. Under the baseline regime, the annual growth rate of the nominal money supply is assumed to be zero, and government spending is also set to zero. The parameter for the transaction cost function was chosen to be 1000. In the scenario of a monetary regime change, there is a shift in the daily growth rate of the nominal money supply from 0 to 1%. This shift does not represent a realistic change in real-world regimes. The purpose of this drastic shift is to facilitate comparison of steady-state values across the two regimes, enabling an AI agent to detect differences.[17]

[15] This involves applying backpropagation and gradient descent procedures.

[16] Similar to the critic network, the specific steps of updating ANN's parameters by minimising an objective function involve backpropagation and gradient descent.

[17] When the two steady states have similar reward values, the AI agent struggles to detect changes effectively. This finding highlights the need for further research and experimentation to better understand how the size or extent of differences in reward values affects the agent's ability to recognize and adapt to these changes.

Table 2: Baseline Parameters and Steady State Values

	Regime I	**Regime II**
Storage technology parameter, α	0.4	
Discount factor, β	0.99999	
Endowment, $\bar{y}$	1	
Government spending, g	0	
Transaction cost function parameter, κ	1000	
Exploration	Low - 0.008; High - 0.02	
Monetary policy parameter, δ^M	1.0	1.01
RE velocity, v	0.1	3.148
Consumption, c	1.326	1.322
Storage, s	0.217	0.217
Real balance, m	13.256	0.420

6 Experiments and Results

6.1 Experiments

In this exercise, I study the adaptive behaviors of a representative AI agent when the environment undergoes a drastic change. I investigate the role of exploration in facilitating the adaptive behaviors of AI agents, and how different levels of exploration may shape the beliefs and behaviors of these agents. Following the full algorithm outlined in Section 4.4, I conduct the following experiment.

A representative AI agent is introduced into an economy with a constant nominal money supply, which then undergoes an increase in the nominal money supply. The change in nominal money supply is unknown to the AI agent. The AI agent operates under the initial policy regime for approximately 27 years. Then, without prior announcement, there is a shift to a new regime where the nominal money supply increases daily at a rate of 1%.[18] The AI agents live in the new regime

[18]This shift does not represent a realistic change in real-world regimes. The purpose of this drastic shift is to facilitate comparison of steady-state values across the two regimes, enabling an AI agent to detect differences.

for around 60 years. Designing the regime change as an adjustment to the money supply process is intended to demonstrate the AI agent's ability to adapt to structural changes, and is rooted in the accelerationist debate.

Given its adaptive and explorative nature, the AI agent is expected to adjust to a monetary policy regime change by adapting its consumption and liquidity holding decisions. Through this experiment, I aim to examine the agent's adaptability, the extent of its adjustments in beliefs and decisions, and how the level of exploration affects its adaptive behavior during such a significant economic shift.

To examine the role of exploration, I set up two AI agents, one with a high level of exploration and the other with a low level, to start in the same economy under identical initial conditions. I then subject both agents to the same regime change experiment to observe and compare their responses. Additionally, I compare the simulation paths of the AI agents against those of an RE agent in both regimes. I replicate the simulation experiment 50 times and analyze the median results. These results, representing the median performance across the repetitions, provide a stable basis for comparison.

Moreover, the plots, derived from simulations conducted during the testing period in which agents operate without further updates or exploration, illustrate what the AI agents would have done after experiencing a regime change and living under the new regime for 60 years. This means that the beliefs have already been updated after experiencing a regime change. Therefore, direct changes in behavior or aggregate variables can be observed as a result of regime changes in the simulation paths. These plots aim to facilitate the comparison of different agents' behaviors and their comparison with the path of the RE agent.

Through this experiment, I highlight three main findings:

1. The AI agent's capability to adapt its beliefs in response to economic changes is evident in its adjustments to consumption, storage, and demand for real balances. The level of exploration significantly affects this adaptability, which in turn impacts aggregate economic transitions, as shown by the inflation levels following an increase in the nominal money supply. Specifically, after the structural change, inflation rises to 1% in the high-exploration economy, while in the low-exploration economy, it stabilizes around 1.5%.

2. As aggregate inflation increases, the total real resources available for AI agents to allocate across decisions—consumption, storage, and liquidity holdings—decrease. Since both agents maintain similar consumption levels in both regimes, they need to reduce their storage and liquidity holdings in the new regime, where fewer real resources are available. The high-exploration agent reduces its liquidity holdings by 25%, more than the low-exploration agent's 22%. As a result, to maintain its consumption level, the low-exploration agent must also reduce its storage, leading to a decrease in wealth. In the new regime, the low-exploration agent

accumulates 0.6% less wealth than the high-exploration agent.

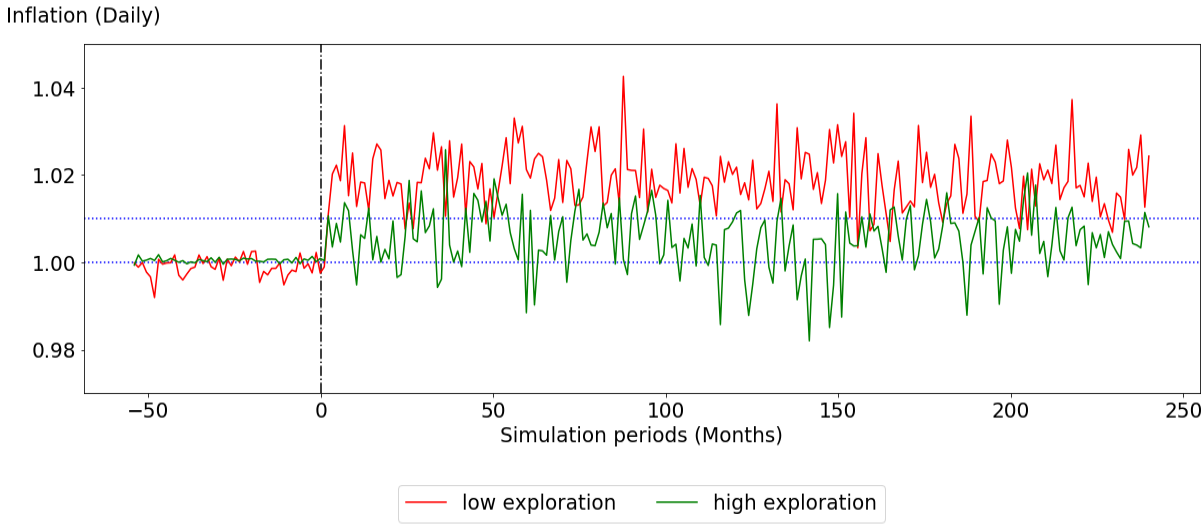
3. Due to their exploration feature, the AI agents' beliefs adjust toward those of the RE agent in both regimes but do not fully align. For example, in the new regime, their liquidity holdings decrease compared to the initial regime, similar to the RE agent, but not by the same amount. However, unlike the RE agent, they demonstrate the ability to adjust and adapt to a regime change.

6.2 Inflation

Inflation rises following an increase in the nominal money supply, but the extent of the increase varies depending on the level of exploration exhibited by the representative AI agent. Before the regime change, inflation in both economies was around 0%, but after an adjustment period, it stabilized at around 1% in one economy and 1.5% in the other, depending on the exploration level of the AI agent. In Figure 2, the red line represents the low-exploration economy, and the green line represents the high-exploration economy. Following the regime shift, inflation in the high-exploration economy becomes volatile before stabilizing at 1%, while in the low-exploration economy, it stabilizes around 1.5%.

In a high-inflation environment, the total real resources available decrease, prompting an adaptive AI agent to adjust its allocations across consumption, storage, and liquidity holding. Examining how the AI agents adjust their consumption, storage, and liquidity holding decisions in response to this regime change can provide better insight into their adaptive behaviors. In particular, will the AI agents reduce their consumption, storage, and/or liquidity holdings to adjust to the reduced overall real resources? How would their behaviors differ based on their exploration levels?

Figure 2: Inflation

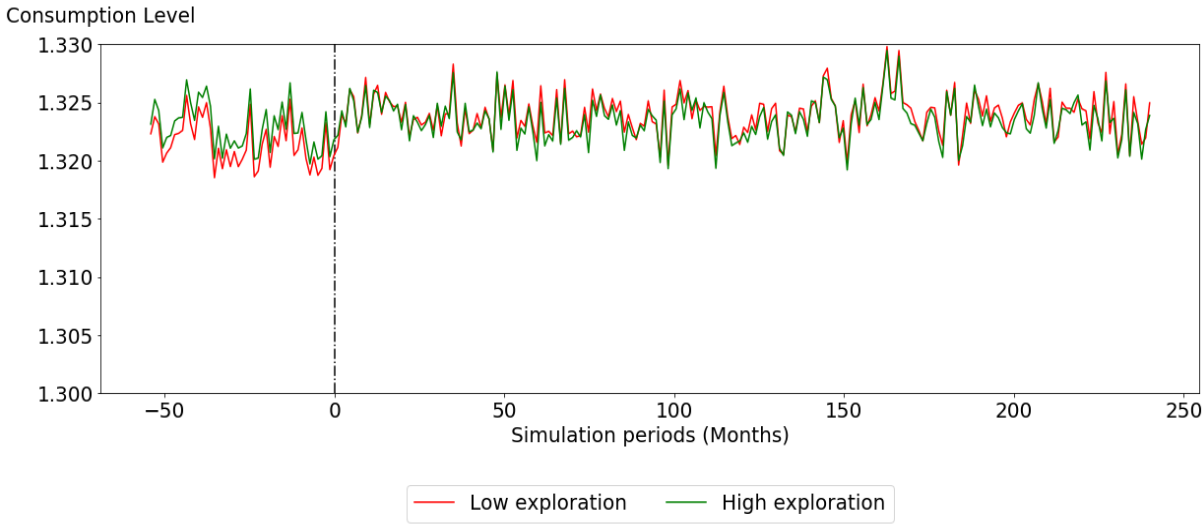


Notes:
1. The x-axis represents simulation periods in months.
2. The y-axis plots daily inflation, where inflation is defined as current period price level divided by the previous period price level.
3. The horizontal dashed lines mark the initial steady state inflation at 0% and the new steady state inflation at 1%, respectively.
4. A vertical dashed line at period 0 represents the occurrence of the regime change.
5. The red line represents the variable of interest for the low-exploration AI agent, and the green line represents that for the high-exploration AI agent.

6.3 Consumption, Storage, Liquidity Holding, and Wealth

Both AI agents exhibit consumption smoothing behavior and reach similar levels of consumption across both regimes, demonstrating their adaptability to structural changes in the economy. Figure 3 illustrates this behavior, beginning 50 months before the regime shift, with period 0 marking the transition. The plots show that consumption remains stable even after the economy shifts to a high-inflation state. This smoothing behavior indicates that both agents achieve similar reward levels before and after the regime change. To sustain their consumption levels despite a reduction in aggregate real resources in a high-inflation environment, the agents must decrease either storage, liquidity holdings, or both.

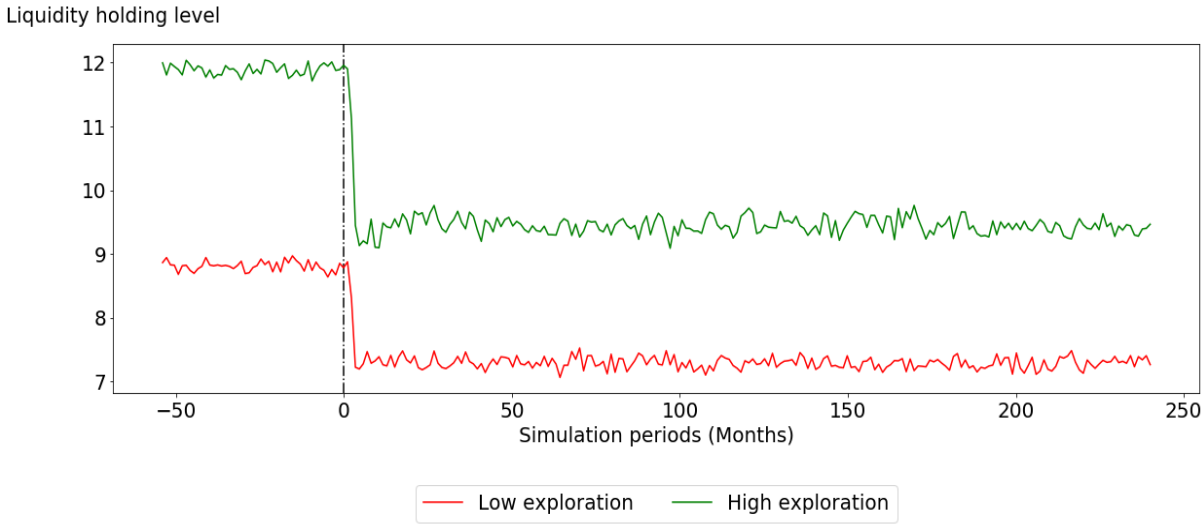
Figure 3: Consumption of AI Agents during a Regime Change



Notes:
1. The x-axis depicts simulation periods in months.
2. The y-axis represents the consumption level.
3. A vertical dashed line at period 0 indicates the occurrence of the regime change. This regime change involves an acceleration of the nominal money supply from a 0% to a 1% daily growth rate.
4. The red line illustrates the variable of interest for the low-exploration AI agent, while the green line represents the same variable for the high-exploration AI agent.

Both agents reduce their liquidity holdings after transitioning from a low- to high-inflation environment to support their desired consumption levels, with the high-exploration agent reducing more than the low-exploration agent. Holding liquidity helps reduce transaction costs when consuming. In a high-inflation regime, real resources are lower than in a low-inflation regime, making it less desirable to allocate the same amount to real balances. Through exploration, AI agents adapt their beliefs and reduce their liquidity holdings to function effectively in the new regime. The high-exploration agent reduces its liquidity holdings by 25% to adjust to the new regime, while the low-exploration agent decreases its liquidity holdings by around 21% following the regime change.

Figure 4: Real balance



Notes:
1. The x-axis represents simulation periods in months.
2. The y-axis displays the level of liquidity holding by the AI agents.
3. A vertical dashed line at period 0 signifies the occurrence of the regime change. This regime change involves an acceleration of the nominal money supply from a 0% to a 1% annual growth rate.
4. The red line represents the variable of interest for the low-exploration AI agent, while the green line depicts the same variable for the high-exploration AI agent.

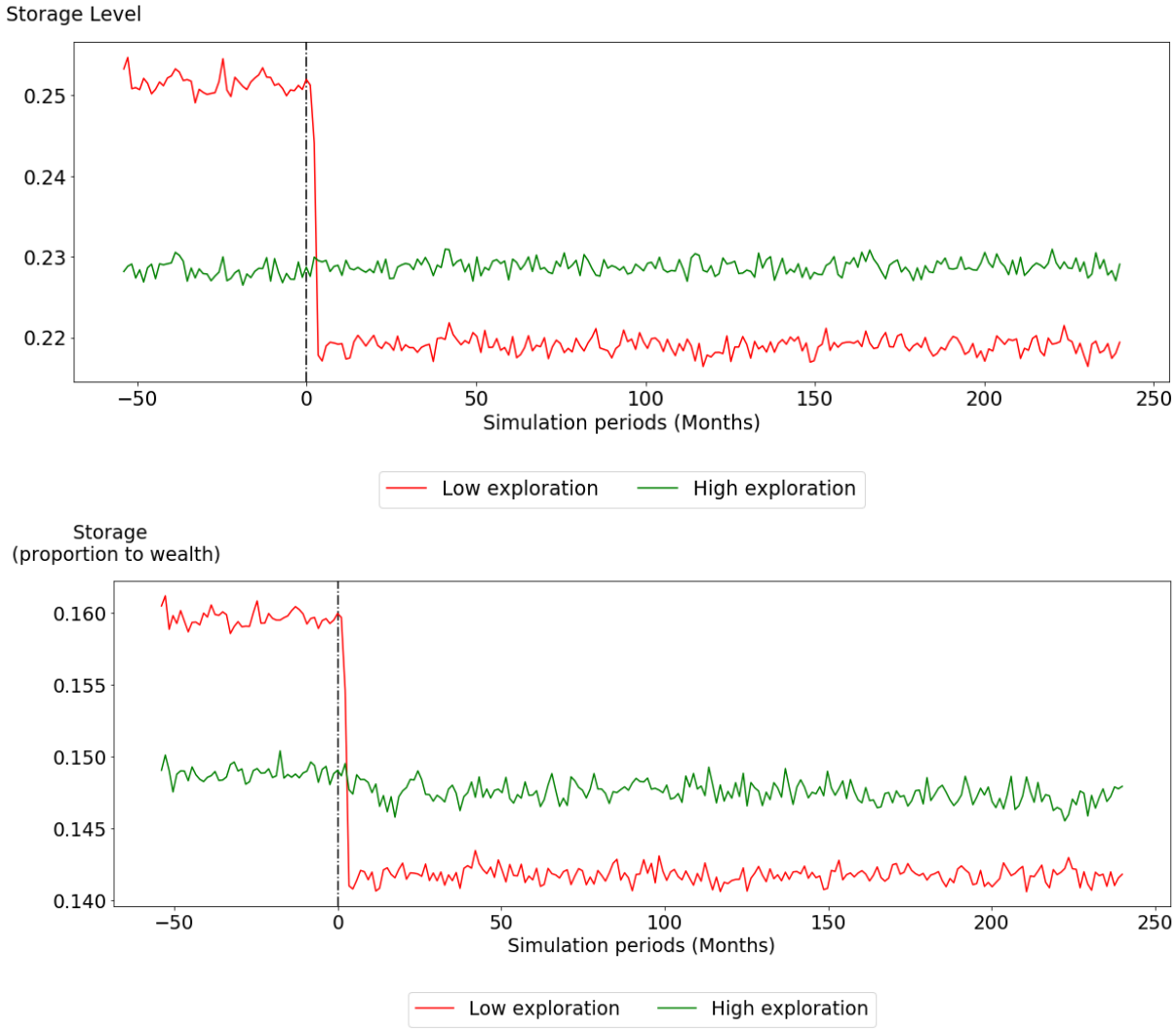
After the regime change, the high-exploration agent sustains its storage and wealth, while the low-exploration agent, reducing liquidity holdings less, must lower its storage to maintain pre-change consumption levels. As shown in Figure 5, the high-exploration agent maintains a storage level of 0.23, while the low-exploration agent maintains a storage level of 0.25. After the regime shift, the low-exploration agent reduces its storage level by 12%, which accounts for around 14% of its wealth, compared to 16% before the regime shift. This reduction in storage contributes to the overall decline in aggregate wealth following the regime change.

The high-exploration agent maintains a consistent storage level, keeping its wealth largely unchanged, while the low-exploration agent's wealth decreases after the regime change, resulting in a 6% difference between the two. Figure 6 shows the wealth levels for AI agents with different exploration levels. Wealth at period t is defined as $y_t + s_t^{\alpha}$, where lower storage means less wealth for the same income, y_t. Both agents start with the same initial wealth (Figure 6, bottom panel). Although their wealth levels differ in the initial regime due to varying exploration,[19] the low-exploration

[19]Since the focus of this paper is on the adaptability of AI agents to regime changes, I will not discuss the variations in outcomes during the initial regime due to different levels of exploration. For more information on this, please refer to Shi (2021).

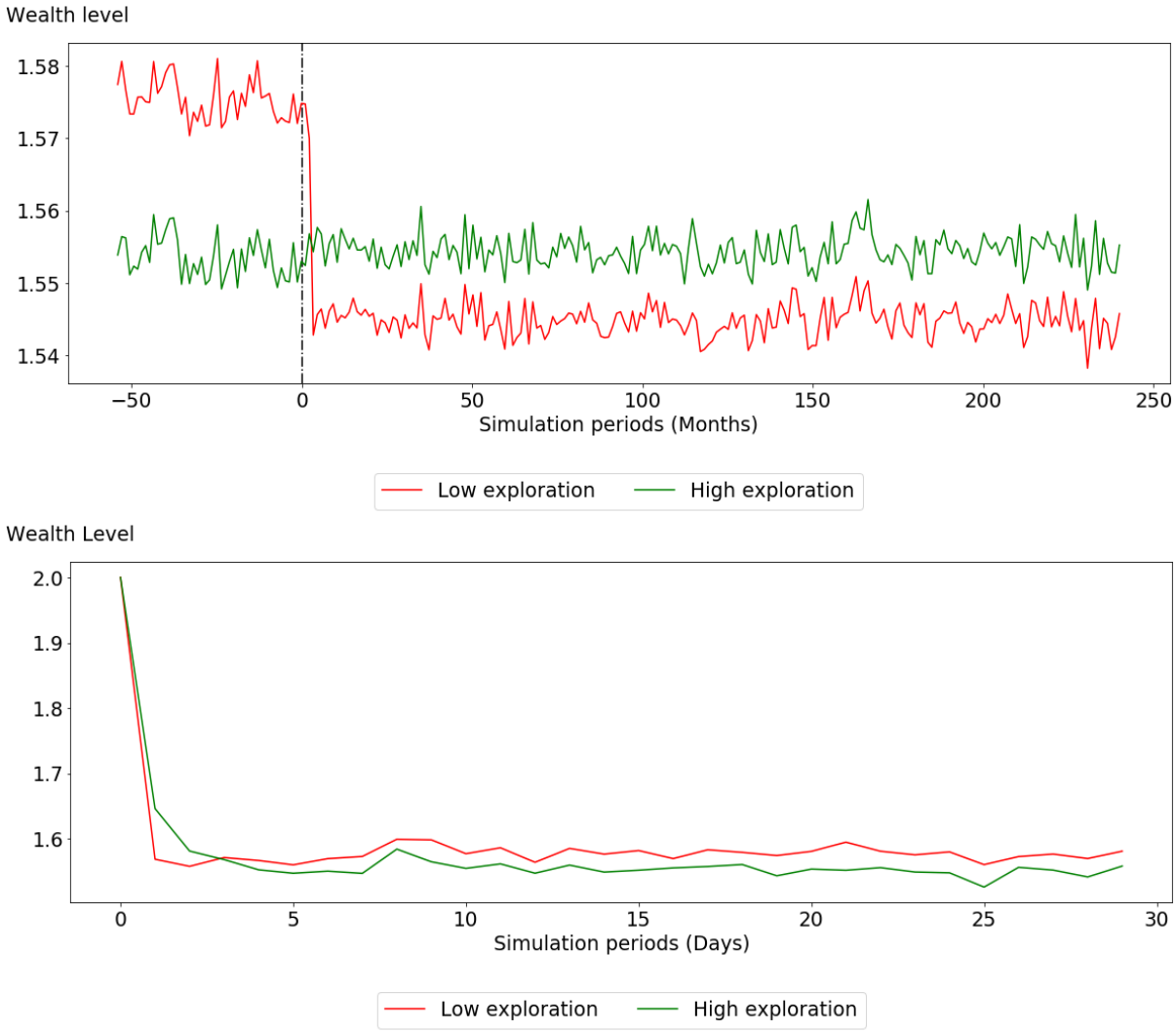
agent's significant storage reduction during the regime change causes its wealth to drop, leaving the high-exploration agent's wealth around 0.6% higher.

Figure 5: Storage of AI Agents during a Regime Change



Notes:
1. The x-axis depicts simulation periods in months.
2. The y-axis represents the savings/storage level in the top panel, and the proportion of savings to wealth in the bottom panel, where wealth at period t is defined as $y_t + s_t^{\alpha}$, encompassing both income and previous storage.
3. A vertical dashed line at period 0 indicates the occurrence of the regime change. This regime change involves an acceleration of the nominal money supply from a 0% to a 1% daily growth rate.
4. The red line illustrates the variable of interest for the low-exploration AI agent, while the green line represents the same variable for the high-exploration AI agent.

Figure 6: Wealth



Notes:
1. The x-axis represents simulation periods in months in the top panel, and days in the bottom panel.
2. The top panel displays the level of wealth in the economy during a regime change, where wealth at period t is defined as $y_t + s_t^{\alpha}$.
3. The initial level of wealth, shown in the bottom panel figure, is identical for both economies, set at an arbitrary value of 2.
4. The red line represents the variable of interest for the low-exploration AI agent, while the green line depicts the same variable for the high-exploration AI agent.
5. A vertical dashed line at period 0 signifies the occurrence of the regime change on the top panel. This regime change involves an acceleration of the nominal money supply from a 0% to a 1% daily growth rate.
6. In the bottom panel, Period 0 represents the initial period of the first regime.

6.4 AI vs RE agent(s)

It's crucial to emphasize that although an RE agent performs well in a stable environment, AI agents adept in adapting to changes within their environment thanks to their exploratory behavior. This adaptability is primarily attributed to the exploration mechanisms inherent in deep RL algorithms, setting them apart from the RE agents. As a result, AI agents are capable of adjusting their behaviors in response to environmental shifts, and their beliefs never fully align with those of an RE agent, sidestepping the limitations posed by the Lucas Critique.

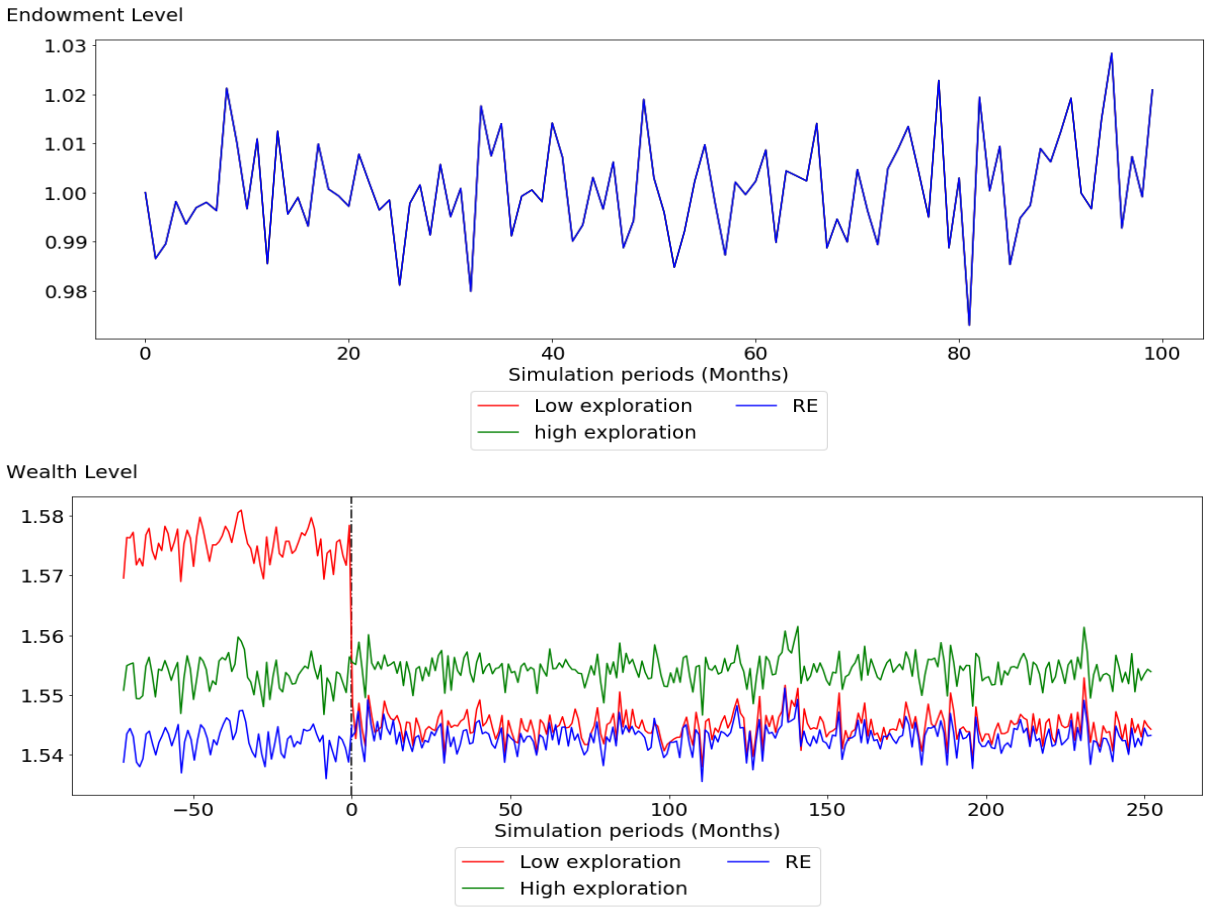
To highlight this distinction, I compare the behavior of an RE agent across the two regimes with AI agents. I solve for the RE equilibrium in each regime separately and then combine the two simulations, labeling this sequence as the RE agent's behavior in Figures 7 and 8.[20] Since the RE agent is constrained by a fixed belief in each regime, this approach avoids the debate over whether the RE agent is aware of the regime change or if it occurs unexpectedly. Additionally, I include the performance of AI agents with both low and high exploration levels, as introduced in the previous section. This comparison highlights the differences between the AI agents and the RE agent. The top panel of Figure 7 depicts the endowment process for both AI and RE agents, illustrating that they operate in an identical environment with the same shock process. As all agents operate under the same conditions, only one line is observable in the figure.

Across both regimes, the consumption levels of all agents are similar but not identical (Figure 8), with AI agents tending to save more than the RE agent. The RE agent reduces its consumption modestly by around 0.3% after the regime shift, while both AI agents seem to consume similar amounts after the change. The high-exploration agent saves about 4% more than the RE agent, while the low-exploration agent saves around 12% more in the initial regime, dropping to just 0.03% more in the new regime. One possible explanation for the higher savings observed in AI agents is that they save extra to allow for more exploration in the future when needed.

Comparing the simulation paths of AI agents with those of the RE agent shows that AI agents adapt their decisions toward the RE agent, though the magnitude of their behavior differs, influenced by their levels of exploration. AI agents' liquidity holdings drop by 22% to 25%, depending on their levels of exploration, while the RE agent's liquidity holdings decrease by around 90%. Additionally, as AI agents save more than the RE agent, they accumulate more wealth. In the new regime, the low-exploration agent accumulates around 0.1% more wealth than the RE agent, while the high-exploration agent accumulates around 0.7% more.

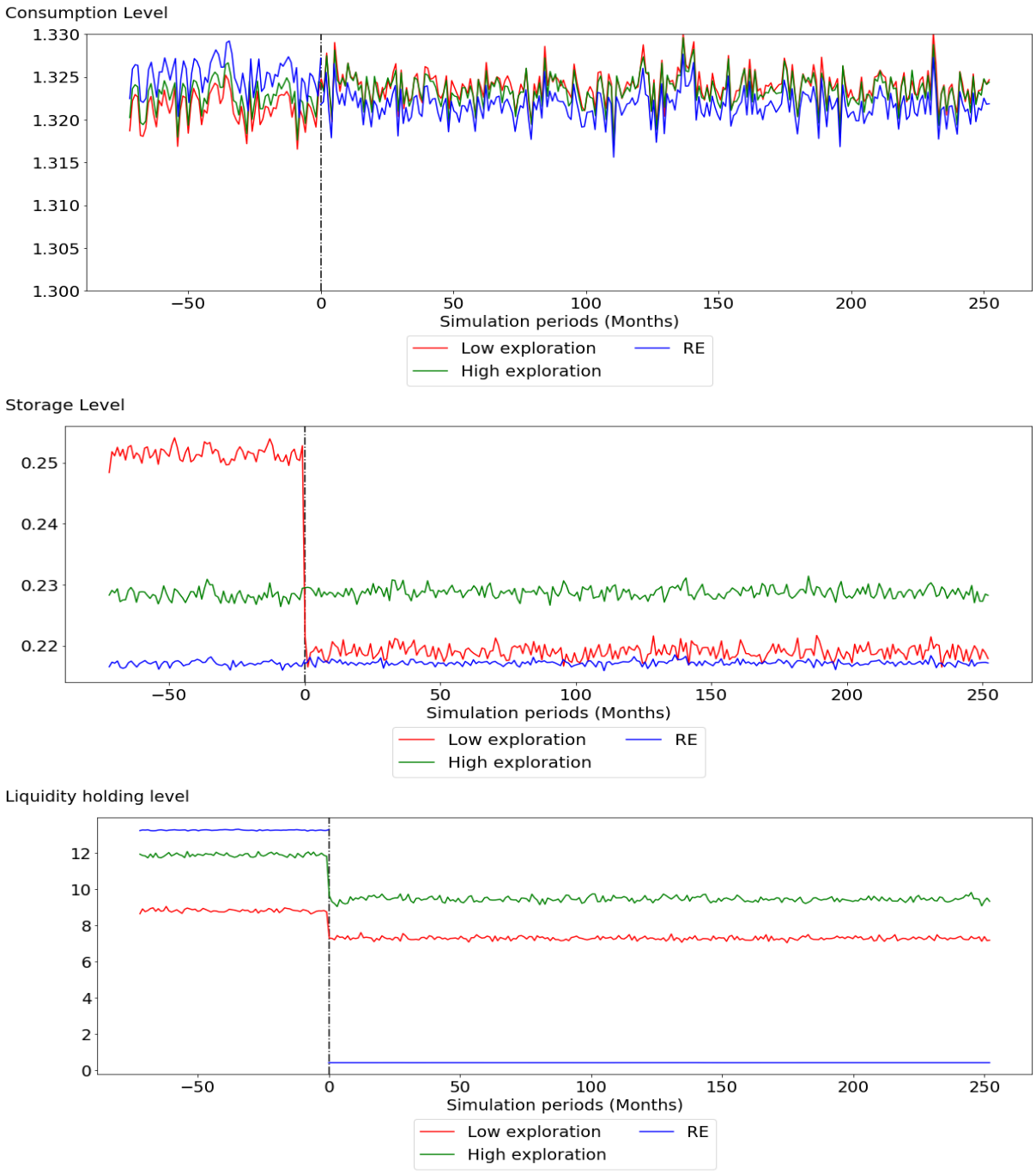
[20] The RE equilibrium is solved using the methodology outlined in Schmitt-Grohe and Uribe (2004), utilizing a second-order approximation of the policy function.

Figure 7: AI vs RE Agents Before and After a Regime Change



Notes:
1. The x-axis depicts simulation periods in months.
2. The y-axis represents endowment level on the top panel, and wealth on the bottom panel.
3. A vertical dashed line at period 0 indicates the occurrence of the regime change. This regime change involves an acceleration of the nominal money supply from a 0% to a 1% daily growth rate.

Figure 8: AI vs RE Agents Before and After a Regime Change



Notes:

1. The x-axis depicts simulation periods in months.
2. The y-axis represents consumption level, storage level, and liquidity holding level, from the top panel to the bottom panel, respectively.
3. A vertical dashed line at period 0 indicates the occurrence of the regime change. This regime change involves an acceleration of the nominal money supply from a 0% to a 1% daily growth rate.

6.5 Discussion

The findings in this exercise demonstrate that AI agents have the ability to adapt their beliefs in response to an accelerating money supply, which has implications on economic transition dynamics and welfare outcomes. The representative AI agent changes its behaviors through interacting with the economic environment. The AI agent makes explorative decisions and observes the associated reward signal. When the current decision-making strategy no longer generates high rewards, the agent adjusts its policy to improve long-term rewards. This aligns with the findings of Cavallo et al. (2017), which demonstrate that private agents are more likely to adjust their inflation expectations in response to price changes in supermarkets rather than relying solely on observed inflation statistics. The direct impact of supermarket price changes on consumers' welfare makes it a more influential signal in shaping consumers' inflation expectations.

It is important to emphasize that, due to their exploratory nature, AI agents do not fully converge to steady-state values, offering a perspective that complements RE models. While exploration entails a trade-off with achieving RE equilibrium, it aligns with aspects of real human behavior, where individuals do not instantly adapt optimally to environmental changes, but instead, learn and adjust through exploration over time. By integrating exploratory behavior and learning from experience, the AI framework, implemented in macroeconomic models, offers an interdisciplinary approach to modeling adaptive economic agents undergoing structural changes. This approach lays the foundation for developing practical models that can be used for policy experimentation and analysis, bridging the gap between theoretical insights and real-world applications.

Nonetheless, implementing deep RL algorithms in policy settings presents several limitations and opportunities for further exploration. A few key points to consider are:

- Slow Learning: Deep RL algorithms have faced criticism for their relatively slow learning and updating processes. The gradual nature of learning in these algorithms may impede AI agents' ability to attain close to the optimal belief within a single generation, especially in complex environments. As explained by Botvinick et al. (2019), the slow learning primarily results from incremental parameter adjustment and weak inductive bias within the algorithm. However, given the rapid evolution of this field, many new deep RL algorithms have been proposed to mitigate this issue and accelerate the learning process. For instance, inspired by Gershman and Daw (2017), deep RL algorithms incorporating episodic memory are under development, enabling AI agents to learn from the experiences gathered by others and potentially reducing the number of required simulation periods. This development highlights the importance of ensuring that adaptive economic agents have access to relevant experiences, enabling them to better adapt to welfare-enhancing changes.
- Utility/Reward Design: Implementing the regime change experiment presents a significant challenge: designing the change as a minor shift in the nominal money supply makes it difficult

for AI agents to notice and adjust. This challenge arises primarily because the change in rewards between the two regimes, while present, is relatively small. Future studies should focus more on improving this reward signal mechanism, ensuring it effectively signals AI agents to adjust their beliefs. This aspect is highly pertinent to policy. Policymakers seeking to accelerate economic transitions need to understand the reward signals that effectively motivate adaptive agents to change their behavior.

- Estimation: Incorporating the exploration mechanism helps integrate theoretical human-like behaviors, leading to noticeable differences between agents' behaviors. However, a crucial unaddressed aspect is aligning parameters with real data or actual human behaviors. Addressing this would make the quantitative implications of exploration more tangible and greatly enhance their applicability in policy analysis.

7 Conclusion

In this paper, I integrated an AI framework into a simple dynamic stochastic general equilibrium model to capture adaptive behaviors of a representative economic agent, drawing on insights from neuroscience and psychology. Using a deep RL approach, the model demonstrated how agents explore alternatives, balance competing objectives, and adjust strategies in response to changes in the nominal money supply. The primary aim was to illustrate the effectiveness of the deep RL framework within a scalable general equilibrium model, showcasing its ability to model an artificial agent that adapts to structural changes through exploration. This work contributes to the existing literature by offering a structured framework and deeper insights into how agents with bounded rationality respond to structural shifts.

Both AI agents adjusted their behavior after a structural change, though they employed different allocation strategies, resulting in distinct outcomes in aggregate inflation and welfare. Specifically, after the increase in nominal money supply, inflation rises to 1% in the high-exploration economy, while in the low-exploration economy, it stabilizes around 1.5%. Exploration not only affects the magnitude of the decisions made by AI agents but also determines which specific decisions are adjusted during a regime shift. The low-exploration AI agent modified both its storage position and liquidity-holding decisions, while the high-exploration AI agent focused primarily on adjusting its liquidity-holding decision. As a result of differences in storage adjustments, the low-exploration agent accumulated 0.6% less wealth than the high-exploration agent in the new regime.

I demonstrated that while the agents' beliefs adjust toward those of an RE agent, they do not fully align due to their exploratory nature. However, this suboptimal belief (measured by comparison with the RE belief) is offset by the AI agents' ability to adapt to regime changes, resembling how humans learn and make decisions in real life.

The findings from this study underscore the utility of the deep RL framework as a valuable tool for modeling artificial agents within an economy undergoing structural changes, while also capturing differences in how economies transition to new states and the resulting impact on overall welfare. The incorporation of the exploration mechanism provided a theoretical, human-like approach to learning and adapting to changes. To enhance the reliability of quantitative predictions for policymakers, future research should focus on anchoring the exploration parameter through real data or actual human behaviors.

References

M. Airaudo and I. Hajdini. Consistent expectations equilibria in markov regime switching models and inflation dynamics. *International Economic Review*, n/a(n/a), 2021. doi: https://doi.org/10.1111/iere.12529. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/iere.12529.

J. Ashwin, P. Beaudry, and M. Ellison. The unattractiveness of indeterminate dynamic equilibria. 2021.

L. Ball, N. Gregory Mankiw, and R. Reis. Monetary policy for inattentive economies. *Journal of Monetary Economics*, 52(4):703–725, May 2005. URL https://ideas.repec.org/a/eee/moneco/v52y2005i4p703-725.html.

D. Bassett and M. Mattar. A network neuroscience of human learning: Potential to inform quantitative theories of brain and behavior. *Trends Cogn Sci.*, 21(4):250–264, 2017. doi: 10.1016/j.tics.2017.01.010. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5366087/#:~:text=Learning%20can%20be%20accompanied%20by,%2C%20removing%2C%20or%20altering%20associations.

M. Botvinick, S. Ritter, J. Wang, Z. Kurth-Nelson, C. Blundell, and D. Hassabis. Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, 23, 04 2019. doi: 10.1016/j.tics.2019.02.006.

P. Cagan. *The Monetary Dynamics of Hyperinflation*, pages 25–117. University of Chicago Press, 1956.

A. Cavallo, G. Cruces, and R. Perez-Truglia. Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, 9 (3):1–35, July 2017. doi: 10.1257/mac.20150147. URL https://www.aeaweb.org/articles?id=10.1257/mac.20150147.

M. Chen, A. Joseph, M. Kumhof, X. Pan, R. Shi, and X. Zhou. Deep reinforcement learning in a monetary model. *Rebuilding Macroeconomics Working Paper Series*, (58), 2021.

M. Curry, A. Trott, S. Phade, Y. Bai, and S. Zheng. Analyzing micro-founded general equilibrium models with many agents using deep reinforcement learning, 2022. URL https://arxiv.org/abs/2201.01163.

G. Evans and S. Honkapohja. Learning dynamics. In J. B. Taylor and M. Woodford, editors, *Handbook of Macroeconomics*, volume 1, Part A, chapter 07, pages 449–542. Elsevier, 1 edition, 1999. URL https://EconPapers.repec.org/RePEc:eee:macchp:1-07.

M. Friedman. *A Theory of the Consumption Function*. Princeton University Press, 1957. https://www.nber.org/books-and-chapters/theory-consumption-function.

S. J. Gershman and N. D. Daw. Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, 68(1):101–128, 2017. doi: 10.1146/annurev-psych-122414-033625. URL https://doi.org/10.1146/annurev-psych-122414-033625. PMID: 27618944.

R. Hertwig and I. Erev. The description–experience gap in risky choice. *Trends in Cognitive Sciences*, 13(12):517–523, 2009. ISSN 1364-6613. doi: https://doi.org/10.1016/j.tics.2009.09.004. URL https://www.sciencedirect.com/science/article/pii/S1364661309002125.

E. Hill, M. Bardoscia, and A. Turrell. Solving heterogeneous general equilibrium economic models with deep reinforcement learning, 2021. URL https://arxiv.org/abs/2103.16977.

N. Hinterlang and A. Tänzer. Optimal monetary policy using reinforcement learning. *Deutsche Bundesbank Discussion Paper No. 51/2021*, 2021.

J. M. Keynes. *The General Theory of Employment, Interest and Money*. Macmillan, 1936. 14th edition, 1973.

A. Kuriksha. An economy of neural networks: Learning from heterogeneous experiences. *arXiv preprint arXiv:2110.11582*, 2021.

T. Lillicrap, J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. *arXiv e-prints*, (arXiv), 2015.

R. Lucas. Expectations and the Neutrality of Money. *Journal of Economic Theory*, 4:103–124, 1972.

R. E. Lucas. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46, 1976. ISSN 0167-2231. doi: https://doi.org/10.1016/S0167-2231(76)80003-6. URL https://www.sciencedirect.com/science/article/pii/S0167223176800036.

U. Malmendier and S. Nagel. Learning from Inflation Experiences *. *The Quarterly Journal of Economics*, 131(1):53–87, 2016. ISSN 0033-5533. doi: 10.1093/qje/qjv037. URL https://doi.org/10.1093/qje/qjv037.

N. G. Mankiw and R. Reis. Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics*, 117(4):1295–1328, 2002. URL https://ideas.repec.org/a/oup/qjecon/v117y2002i4p1295-1328..html.

V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing atari with deep reinforcement learning. In *NIPS Deep Learning Workshop*. 2013.

T. Ozden and R. Wouters. Restricted Perceptions, Regime Switches and the Effective Lower Bound. 2021. URL https://www.tolgaozden.net/doc/ozdenwouters2020_draft.pdf.

T. Sargent. *Bounded Rationality in Macroeconomics*. Oxford University Press, 1993.

T. J. Sargent. A note on the "accelerationist" controversy. *Journal of Money, Credit and Banking*, 3 (3):721–725, 1971. ISSN 00222879, 15384616. URL http://www.jstor.org/stable/1991369.

T. J. Sargent and N. Wallace. Rational expectations and the dynamics of hyperinflation. *International Economic Review*, 14(2):328–50, 1973. URL https://EconPapers.repec.org/RePEc:ier:iecrev:v:14:y:1973:i:2:p:328-50.

S. Schmitt-Grohe and M. Uribe. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control*, 28(4): 755–775, 2004. ISSN 0165-1889. doi: https://doi.org/10.1016/S0165-1889(03)00043-5. URL https://www.sciencedirect.com/science/article/pii/S0165188903000435.

F. Schorfheide. Learning and Monetary Policy Shifts. *Review of Economic Dynamics*, 8(2):392–419, April 2005. doi: 10.1016/j.red.2005.01.001. URL https://ideas.repec.org/a/red/issued/v8y2005i2p392-419.html.

W. Schultz, P. Dayan, and P. R. Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997. doi: 10.1126/science.275.5306.1593. URL https://www.science.org/doi/abs/10.1126/science.275.5306.1593.

D. Shah, Q. Xie, and Z. Xu. Stable reinforcement learning with unbounded state space, 2020.

R. A. Shi. Learning from zero: How to make consumption- saving decisions in a stochastic environment with an ai algorithm. *CESifo Working Paper No. 9255*, 2021. https://ssrn.com/abstract=3907739.

R. A. Shi. Deep reinforcement learning and macroeconomic modelling. April 2023. URL https://wrap.warwick.ac.uk/id/eprint/185973/. Unpublished.

C. Sims. A simple model for study of the determination of the price level and the interaction of monetary and fiscal policy. *Economic Theory*, 4(3):381–99, 1994. URL https://EconPapers.repec.org/RePEc:spr:joecth:v:4:y:1994:i:3:p:381-99.

C. A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3): 665–690, 2003. ISSN 0304-3932. doi: https://doi.org/10.1016/S0304-3932(03)00029-1. URL https://www.sciencedirect.com/science/article/pii/S0304393203000291. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

A. Subramanian, S. Chitlangia, and V. Baths. Reinforcement learning and its connections with neuroscience and psychology. *Neural Networks*, 145:271–287, 2022. ISSN 0893-6080.

doi: https://doi.org/10.1016/j.neunet.2021.10.003. URL https://www.sciencedirect.com/science/article/pii/S0893608021003944.

R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2018.

R. S. Sutton and A. G. Barto. Toward a modern theory of adaptive networks: expectation and prediction. *Psychological review*, 88(2):135, 1981.

A. Tohid and R. A. Shi. Deep reinforcement learning: Emerging trends in macroeconomics and future prospects. *IMF Working Paper Series No. 2022/259*, 2022.

M. Woodford. Imperfect Common Knowledge and the Effects of Monetary Policy. NBER Working Papers 8673, National Bureau of Economic Research, Inc, Dec. 2001. URL https://ideas.repec.org/p/nbr/nberwo/8673.html.

Address: 10 Shenton Way, #15-08
MAS Building, Singapore 079117
Website: www.amro-asia.org
Tel: +65 6323 9844
Email: enquiry@amro-asia.org
LinkedIn | Twitter | Facebook | YouTube