Project 1

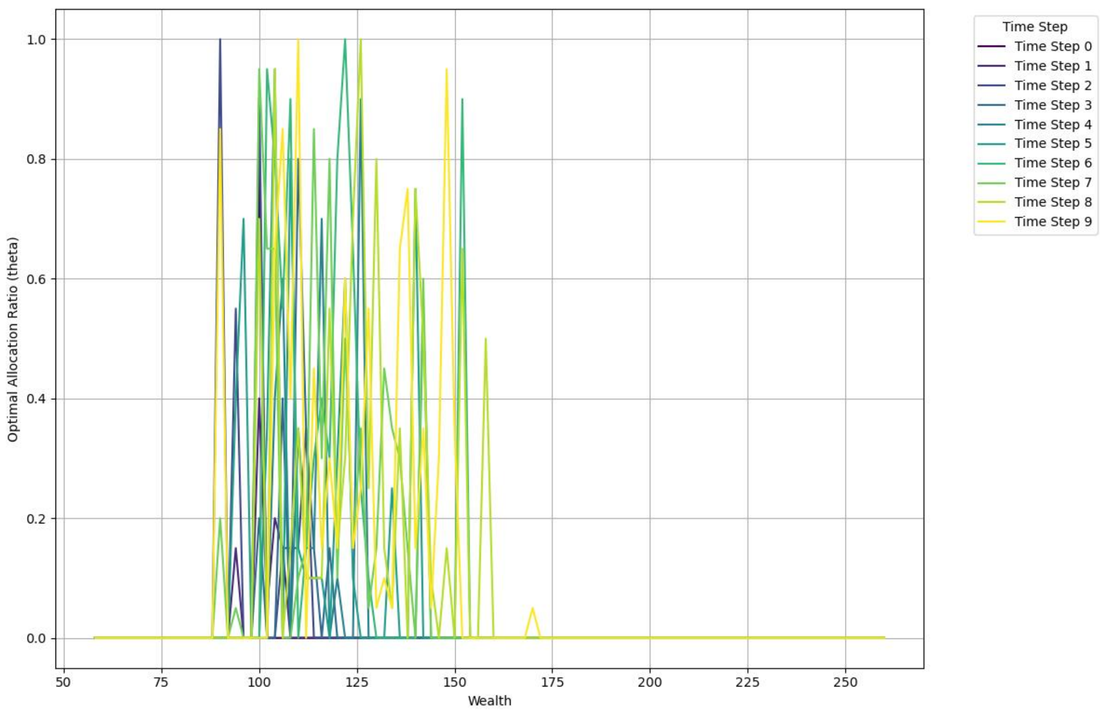
Members:

Deren Zhang: Baseline model design;

Hailong Liu: Heterogeneity and parameter sensitivity analysis.

In this project, we employed the Temporal Difference (TD) method to conduct extensive simulations, deriving an approximate expression for the Q-function and determining the optimal policy. The parameters were set as follows: time horizon T=10, initial wealth W0=100, risky asset parameters a=0.1, b=−0.05, p=0.6, risk-free asset return r=0.02, discount factor gamma = 0.9, learning rate alpha = 0.1, exploration rate epsilon = 0.1, and number of training episodes num_episodes=100,000.
To reduce computational complexity, we calculated the maximum wealth W_max and minimum wealth W_min based on the two assets. Wealth was then discretized into bins with an interval of wealth_bin_interval=2.0. For each step, the allocation to the risky asset was chosen from a range of 0 to 1, with increments of 0.05.

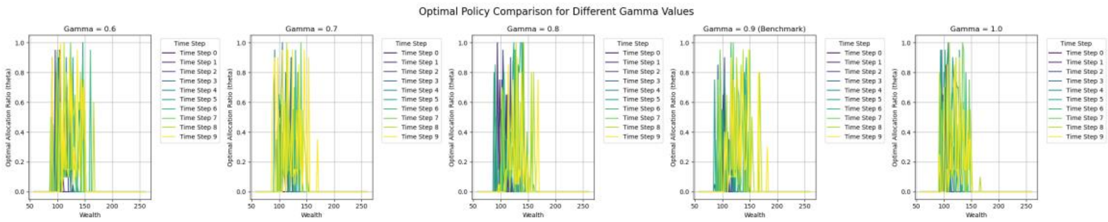
Subsequently, we ran simulations and plotted the optimal allocation to the risky asset against wealth at different time points on a single graph. The results are presented below:



The above describes the baseline model. Below are the sensitivity analysis conducted.

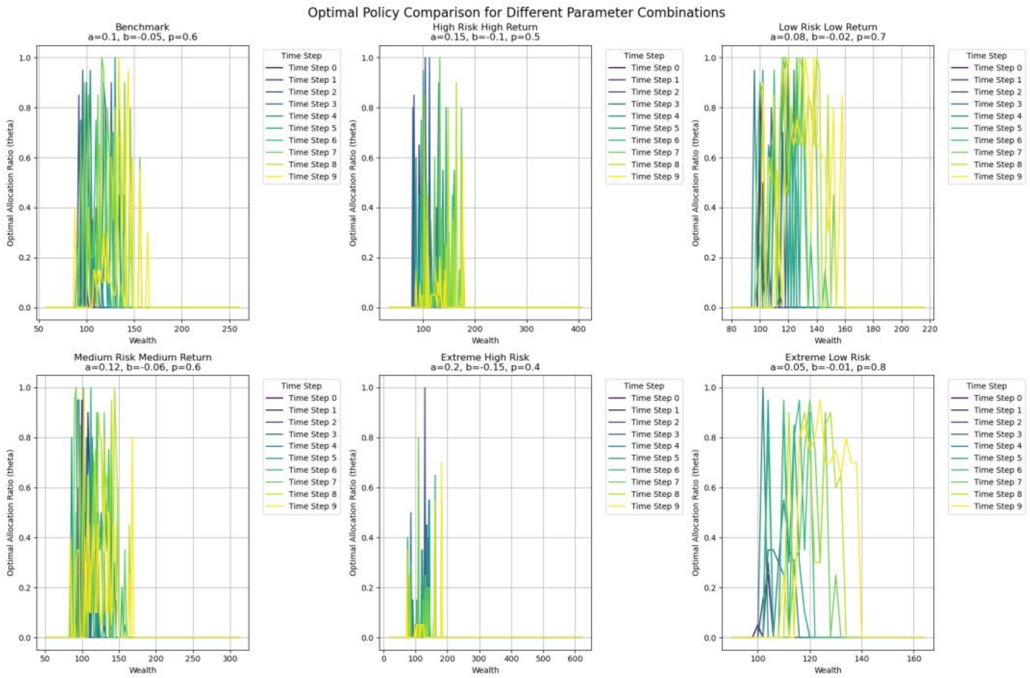
1. Impact of Discount Factor (gamma)

Firstly, adjustments to the discount factor demonstrated negligible effects on policy outcomes. This result could be attributed to the *sparse reward problem*, where rewards are concentrated at the terminal period (T = 10) and are 0 for all other time steps, thereby diminishing the influence of future reward discounting.



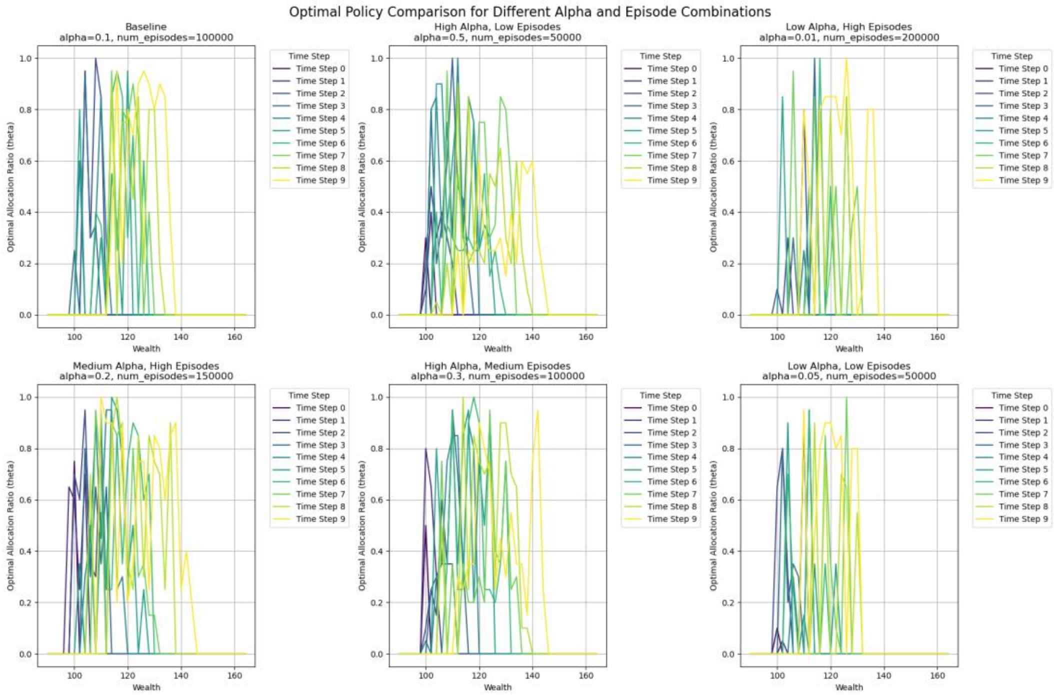
2. Impact of Risky Asset Parameters (a, b, p)

Secondly, adjusting the expected return and success probability of the risky asset had a noticeable impact on optimal allocations. As risk levels increased, allocations to the risky asset decreased, which was in line with expectations. However, an unusual drop in allocations was observed under extremely low-risk conditions. This was likely due to approximation errors caused by the coarse discretization of wealth levels.



3. Impact of Learning Rate (alpha) and Number of Episodes (num_episodes)

Thirdly, the analysis highlighted the importance of properly tuning the learning rate and training episodes. When the learning rate was set too high or too low, it disrupted convergence and led to poor results. Similarly, using too few training episodes resulted in suboptimal policies. Achieving the best outcomes required careful calibration of these parameters, with longer training cycles proving beneficial when feasible.



4. Impact of Exploration Rate (epsilon)

Finally, experiments with fixed exploration rates showed the importance of managing the balance between exploration and exploitation. If the exploration rate was too low, it slowed down policy improvements, while overly high rates prevented the policy from refining effectively. Dynamic strategies, such as gradually reducing the exploration rate over time, produced better results by allowing a smooth transition from exploration to exploitation.