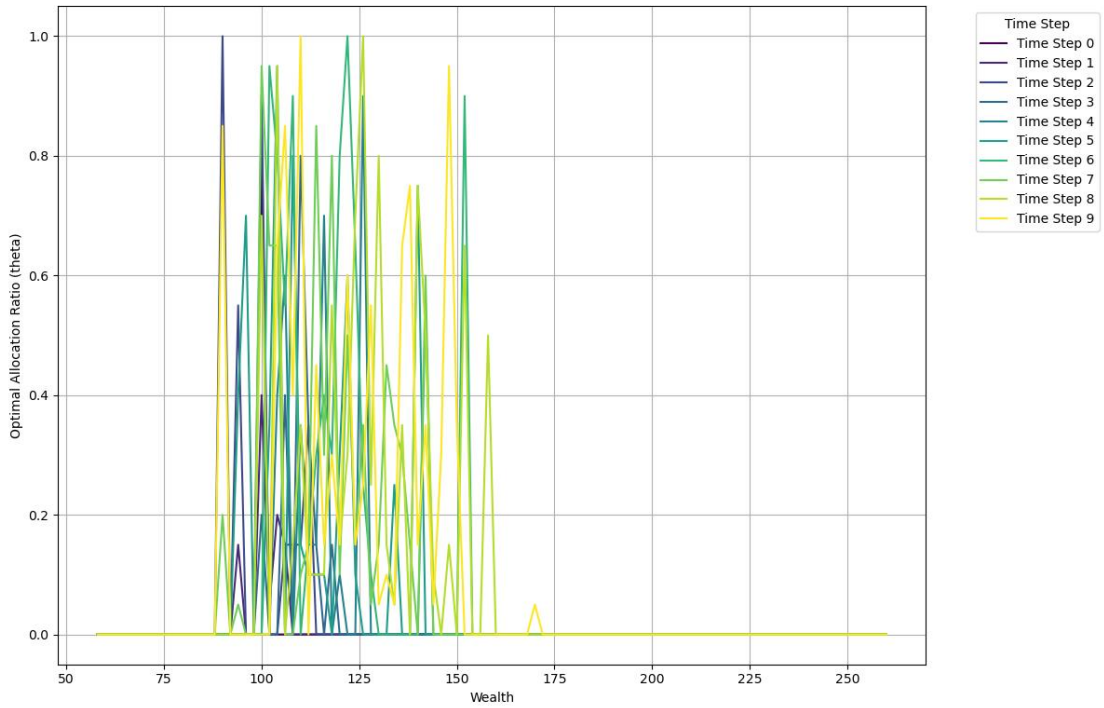
Project 1

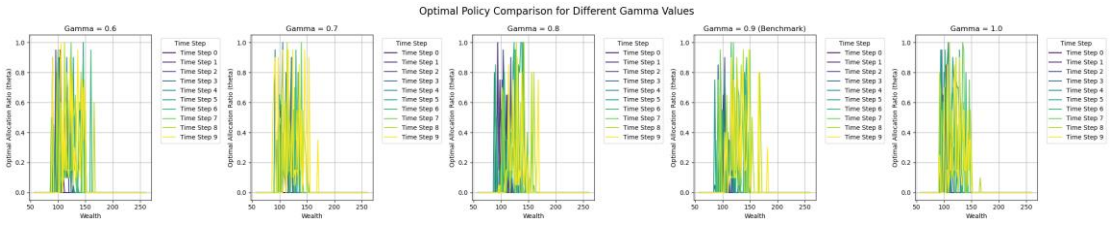
In this project, we employed the Temporal Difference (TD) method to conduct extensive simulations, deriving an approximate expression for the Q-function and determining the optimal policy. The parameters were set as follows: time horizon T=10, initial wealth W0=100, risky asset parameters a=0.1, b=−0.05, 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:



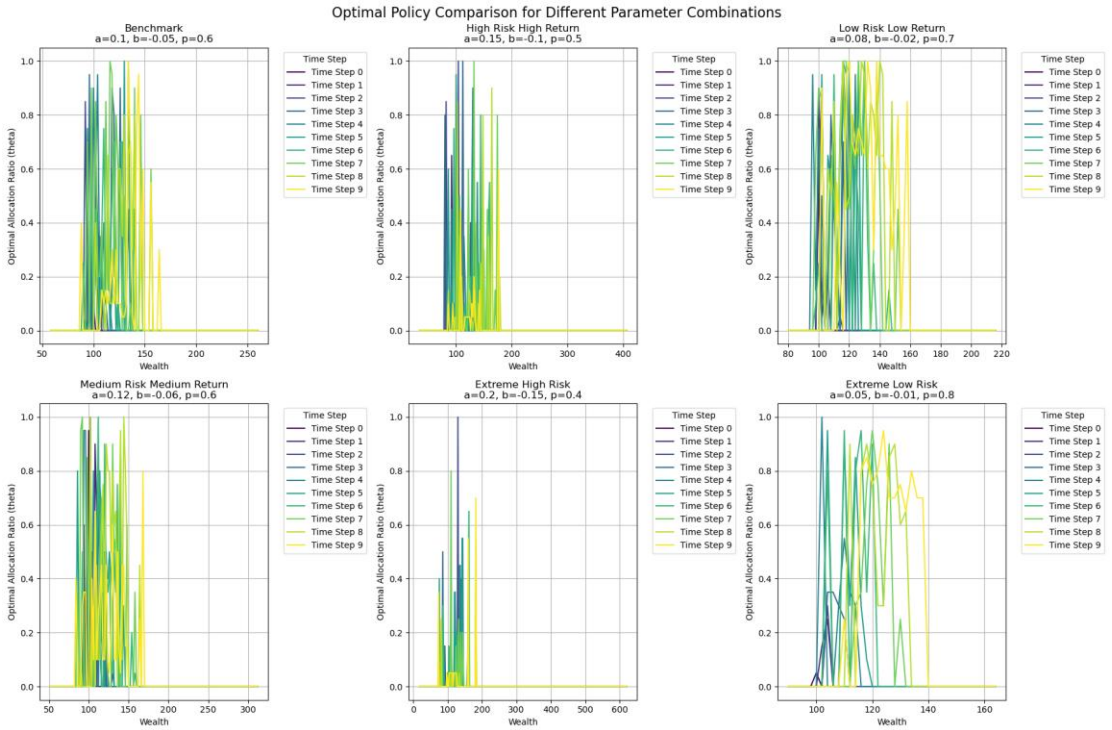
The above describes the baseline model. Next, we conducted a parameter sensitivity analysis.

1. Varying the Discount Factor (gamma)
We first altered the discount factor gamma. The results showed that changes in gamma had no significant impact on the outcome. This can be attributed to the problem's structure, where the reward is positive only at T=10 and zero at all other time steps.



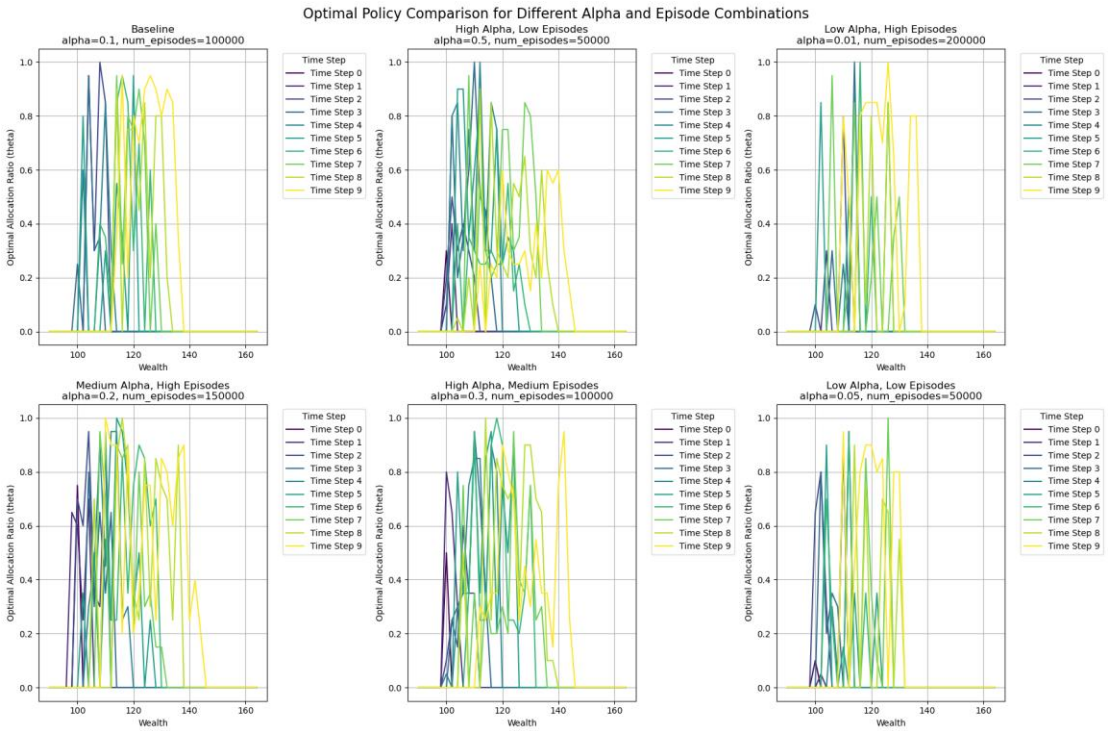
2. Varying Risky Asset Parameters (a, b, p)

Next, we modified the risky asset parameters, testing six distinct styles of risky assets. The results indicated that as the risk of the asset increased, the optimal allocation to the risky asset noticeably decreased. Interestingly, when the risk was extremely low, the allocation to the risky asset also decreased. This anomaly is likely due to the relatively large wealth bin size, which introduced significant approximation errors.



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

We then adjusted the learning rate alpha and the number of training episodes num_episodes. The findings underscored the importance of selecting an appropriate learning rate and sufficient training episodes. Excessively high or low learning rates led to inadequate training. To achieve optimal performance, a suitable learning rate should be chosen, and the number of training episodes should be maximized where feasible.



4. Varying Exploration Rate (epsilon)

Finally, we varied the exploration rate epsilon. The results suggested that when using a fixed exploration rate, an appropriate value is critical—too low a rate resulted in insufficient training. When employing a dynamic exploration rate, strategies such as linear decay, exponential decay, and adaptive exploration all yielded favorable outcomes.