Abstract

Delta hedging is an options strategy that utilizes delta-aiming to reduce the risk associated with price movements in the underlying asset while minimizing trading costs, by taking offsetting long or short positions to hedge an option. Delta hedging is derived by the well-celebrated Black-Scholes Merton (BSM) model, where the pricing of a European call option is given by the equation:

\[ C = S_0 N(d_1) - X e^{-rT} N(d_2) \]

where:
\( S_0 \) is the current stock price,
\( X \) is the strike price,
\( r \) is the risk-free interest rate,
\( \sigma \) is the volatility of the underlying asset,
\( T \) is the time to maturity,
\( N(d_1) \) and \( N(d_2) \) are the cumulative distribution functions of the standard normal distribution.

Our benchmark strategy, delta hedging, is derived by the well-celebrated Black-Scholes Merton (BSM) model, where the pricing of a European call option is given by the equation:

\[ C = S_0 N(d_1) - X e^{-rT} N(d_2) \]

where:
\( S_0 \) is the current stock price,
\( X \) is the strike price,
\( r \) is the risk-free interest rate,
\( \sigma \) is the volatility of the underlying asset,
\( T \) is the time to maturity,
\( N(d_1) \) and \( N(d_2) \) are the cumulative distribution functions of the standard normal distribution.

We employed Deep Reinforcement Learning (DRL) to address this hedging problem in a realistic setting, including discrete time trading with high level of market friction. First, we implemented a simulation environment that simulated stock movements and option pricing by OpenAl Gym. Second, we utilized multiple DRL methods including Deep Q-Learning (DQL), DQL with Preserve Output Precisely and Adaptive Scaling Target (Pop-art) implementation, and Proximal Policy Optimization (PPO) to build agents that can learn how to optimally hedge an option. Third, we evaluated the agent performance in terms of accumulative reward, volatility, trading cost, and profit and loss simulation of our agents and the baseline delta hedging policy. We are able to show that PPO has the best performance among all other DRL algorithms. Moreover, PPO has significantly shorter training time and generates more financially sensible policy than other DRL methods.

Finance Background

An option is a financial derivative that gives the buyer the right to buy or sell a number of corresponding underlying assets at a set price point on a set future date. Our benchmark strategy, delta hedging, is derived by the well-celebrated Black-Scholes Merton (BSM) model, where the pricing of a European call option is given by the equation:

\[ C = S_0 N(d_1) - X e^{-rT} N(d_2) \]

where:
\( S_0 \) is the current stock price,
\( X \) is the strike price,
\( r \) is the risk-free interest rate,
\( \sigma \) is the volatility of the underlying asset,
\( T \) is the time to maturity,
\( N(d_1) \) and \( N(d_2) \) are the cumulative distribution functions of the standard normal distribution.

We integrated trading cost to reward by the following equation:

\[ R_t = R_{t-1} + K \frac{1}{2} (\delta W_t)^2 \]

where:
\( R_t \) is the accumulative reward at time \( t \),
\( K \) is the cost multiplier,
\( \delta W_t \) is the change in the stock price at time \( t \).

We will solve the problem in a more realistic setting where we perform discrete hedging with transaction cost in a market with friction.

Conclusion & Discussion

Training time and Convergence

DQL and PPOs with reward clipping will have crashed reward if training takes too long. DQL with Pop-art can fix this issue. In general, PPO has the fastest convergence speed than all other methods.

Accumulative Reward

All DRL agents have a similar policy to baseline delta hedging.

Total Cost and Volatility

Non-cost case (cost multiplier = 0)

- All DRL agents find more optimal strategy as the average realized vol are much lower compared to baseline delta, but slightly large than zero as financially, given discrete time trading, DRL agents tend to be off a bit in between hedging time.
- High-cost case (cost multiplier = 5)

- Delta trades too much inducing higher cost compared to all DRL agents.
- All DRL agents realize much lower average cost while maintaining the hedge, showing their capability balancing between trading error and costs.
- Overall PPO achieves better performance in terms of its lower average cost at 12.76% compared to both DQL and DQL with Pop-art, at the sacrifice of slightly higher volatility of total P&L, representing its more cost-conscious decision at trade-off between costs and trading errors.

P&L

The mean of Delta Policy’s P&L is significantly smaller than zero in both non-cost and high-cost cases. All DRL agents outperform Delta as their t-statistic of P&L are much more often close to zero and insignificant. DQL with Pop-art performs slightly better in high-cost case compared to other DRL agents, as it achieves significantly positive t-statistics of total P&L.

Policy

All agents trade less when cost is implemented and the number of random actions (individual data deviating from the piecewise segments) decreases. PPO trades more conservatively than both DQN agents. There is an apparent effect of Pop-art on DQL: less random actions are made, and the decisions are closer to the benchmark decisions (dashed lines, derived from delta hedging).

Acknowledgements

It is our pleasure to have our wonderful advisor Petter Nkilmo leading and supporting us throughout the project.