Deep Hedging of Derivatives Using Reinforcement Learning

Presented by: Erik Huebner

Outline

- Introduction
- Reinforcement Learning
- Application to Hedging
- Experiments
- Conclusions

Introduction

What is Hedging?

Hedging is a strategy used by investors to mitigate the risk associated with investing

Investment Terminology

Two types of positions:

- Long
- Short

Two basic types of derivatives:

- Put
- Call

Example: Long Put

You buy 10 shares of a company at $10 per share for $100

You face unlimited potential profit, but also risk losing $100

You might buy a put option for $25 with a strike price of $5 per share

Options

Options/derivatives can also be used directly for investing

This also carries some risk

Derivatives can be used to hedge previous derivatives

Delta Hedging

- Measured as the ratio between the change in the price of the option and the change in the price of the underlying asset:

$$delta = \frac{\Delta C}{\Delta V_A}$$

- Put options have delta between -1 and 0
- Call options have delta between 0 and 1
- The typical strategy is to reach a "delta neutral" position

Rebalancing

- Because the delta of an option changes over time, a trader's position must be rebalanced
- With no transaction costs, continuous rebalancing is optimal
- With transaction costs, even periodic delta hedging is not optimal
- Optimal strategy depends on past holdings and future states

Reinforcement Learning

Reinforcement Learning Formulation

- Reinforcement learning algorithms attempt to solve an MDP by finding an optimal policy that maximizes the expected value of discounted rewards:

$$\boxed{G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots + \gamma^{T-1} R_T}$$

- Where:
  - T is the horizon date
  - R_t is the cash flow received at time t
  - γ is the discount rate

Reinforcement Learning Techniques

- Monte Carlo
- Temporal Difference
- Q-Learning
- Policy Update
- Deep Q-Learning
- Deterministic Policy Gradient

Application to Hedging

Application to Hedging

Assumed that the trader is in a short position in a call option

- Trading costs are proportional to volume being bought/sold
- Rebalance position every Δt periods
- Horizon is $n\Delta t$

The state S at time $i\Delta t$ is defined by:

- The holding of the asset at the previous period
- The current asset price
- The time to maturity

Reward Formulation: Accounting P&L

$$R_{i+1} = V_{i+1} - V_i + H_i(S_{i+1} - S_i) - k|S_{i+1}(H_{i+1} - H_i)|$$

- Where:
  - S_i is the asset price at the beginning of period i
  - H_i is the holding between time i and i+1
  - k is the proportion of trading cost
  - V_i is the value of the option at beginning of period i

Reward Formulation: Cash Flows

$$R_{i+1} = S_{i+1}(H_i - H_{i+1}) - k|S_{i+1}(H_{i+1} - H_i)|$$

- Where:
  - S_i is the asset price at the beginning of period i
  - H_i is the holding between time i and i+1
  - k is the proportion of trading cost

Approach

- Objective function:

$$Y(t) = E(C_t) + c\sqrt{E(C_t^2) - E(C_t)^2}$$

- C_t is the cost of hedging (negative returns) and we seek to minimize this objective function
- Deep Deterministic Policy Gradient algorithm is used with replay buffer

Approach

- The greedy action, a, minimizes F:

$$F(S_t, a) = Q_1(S_t, a) + c\sqrt{Q_2(S_t, a) - Q_1(S_t, a)^2}$$

- Two Q-Functions are used:
  - Q_1 estimates the expected cost
  - Q_2 estimates the expected square of the cost

Approach

- Loss function for Q_1, parameterized by w_1

$$\left(R_{t+1} + \gamma Q_1\left(S_{t+1}, \pi(S_{t+1})\right) - Q_1(S_t, A_t; w_1)\right)^2$$

- Loss function for Q_2, parameterized by w_2

$$\left(R_{t+1}^2 + \gamma^2 Q_2\left(S_{t+1}, \pi(S_{t+1})\right) + 2\gamma R_{t+1} Q_1(S_{t+1}, \pi(S_{t+1})) - Q_2(S_t, A_t; w_2)\right)^2$$

- Policy update:

$$\theta \leftarrow \theta - \alpha \nabla_\theta F(S_t, \pi(S_t;\ \theta))$$

Experiments

Geometric Brownian Motion

- Assume that the price of the underlying asset, S, follows a geometric Brownian motion:

$$\boxed{dS = \mu S dt + \sigma S dz}$$

- Where:
  - μ is the stock's mean return
  - σ is the stock's **constant** volatility
  - dz is a Wiener process

Results

Rebal Freq	Delta Hedging		RL Optimal Hedging		$Y(0)$ improvement
	Mean Cost	S.D. Cost	Mean Cost	S.D. Cost	
weekly	69%	50%	60%	54%	1.7%
3 days	78%	42%	62%	48%	4.7%
2 days	88%	39%	73%	41%	8.5%
daily	108%	38%	74%	42%	16.6%

Rebal Freq	Delta Hedging		RL Optimal Hedging		$Y(0)$ improvement
	Mean Cost	S.D. Cost	Mean Cost	S.D. Cost	
weekly	55%	31%	44%	38%	0.2%
3 days	63%	28%	46%	32%	10.9%
2 days	72%	27%	50%	29%	16.6%
daily	91%	29%	53%	28%	29.0%

- Mean stock return is 5%
- Volatility is 20%
- c = 1.5
- k = 1%

Stochastic Volatility

- Assume that the price of the underlying asset, S, follows a geometric Brownian motion, but with stochastic volatility:

$$\boxed{dS = \mu S dt + \sigma S dz_1}$$

$$\boxed{d\sigma = v\sigma dz_2}$$

- Where:
  - dz_1 and dz_2 are Weiner processes with constant correlation p
  - v is the volatility of the volatility

Results

Rebal Freq	Bartlett Delta		Practitioner Delta		RL Optimal		$Y(0)$ improv. vs. Bartlett	$Y(0)$ improv. vs. Delta
	Mean	S.D.	Mean	S.D.	Mean	S.D.		
weekly	69%	51%	69%	50%	56%	57%	2.6%	1.8%
3 days	78%	44%	78%	43%	61%	51%	4.5%	3.5%
2 days	88%	41%	88%	40%	62%	52%	6.9%	6.0%
daily	108%	39%	108%	38%	71%	45%	16.7%	15.9%

Rebal Freq	Bartlett Delta		Practitioner Delta		RL Optimal		$Y(0)$ improv. vs. Bartlett	$Y(0)$ improv. vs. Delta
	Mean	S.D.	Mean	S.D.	Mean	S.D.		
weekly	55%	36%	55%	35%	42%	43%	2.5%	0.5%
3 days	64%	33%	64%	32%	48%	39%	7.3%	5.3%
2 days	72%	33%	72%	31%	54%	34%	13.7%	11.9%
daily	91%	35%	91%	33%	46%	38%	27.9%	26.4%

- Mean stock return is 5%
- Volatility is 20%, σ_0 is 20%
- p = -0.4
- v = 0.6

Conclusions

Successes

- Successfully applied reinforcement learning to learn an optimal hedging strategy
- Accounting P&L formulation works better than cash flow formulation
- Combining simple option pricing model with complex asset pricing model yields best results

Estimation of Volatility

- Proposed solution accurately estimates the standard deviation of the cost of hedging
- Demonstrates how standard RL algorithms can be used to estimate first and second non-central moments of the cost distribution for hedging

Extensions

- Allow transaction costs to be stochastic
- Estimate optimal strategy for more exotic options
- Use mixture model to generalize across a set of asset price processes

Thanks for Watching!