Discrete Time Hedging with Trading Cost
A Reinforcement Approach

Yilin Liu, Philip Sun

Investment Presentation:

Presentation Overview

adaptive

Acknowledgements

AlgoSeek

Boston University, Questrom School of Business

- Hao Xing
- Chris Kelliher

adaptive

The Problem

- Adaptive offers personalized downside protection recommendations for customers' investment portfolios. To reduce the sell-side risk, we hedge the option with underlyings.
- Ostensibly, delta hedging is easy to implement. However, it is suboptimal for discrete time trading.
- Our team seeks to develop a discrete time model with AI techniques that takes trading cost into consideration.

Architecture

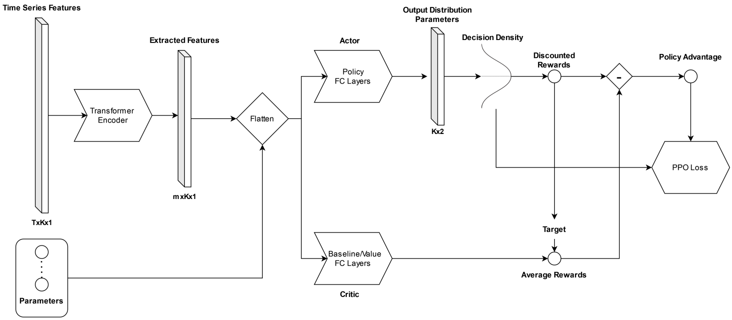

Figure 1: AI Hedging Architecture

Incorporating Time Series Features

- We use transformer models to incorporate time series as input, following are some advantages over recursive frameworks (e.g. RNN, LSTM, GRU):
  - Vanishing gradient problem
  - Computational efficiency

adaptive

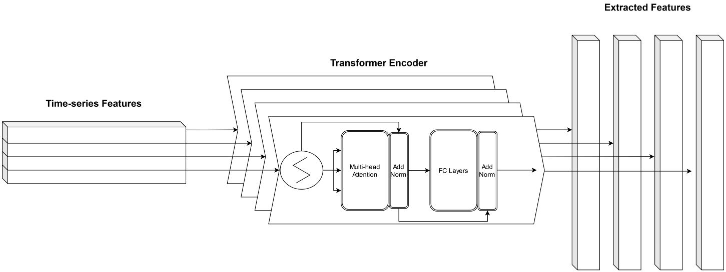

Figure 2: Feature Extraction

adaptive

Positional Encoding

$$P(x_i^k) = \begin{cases} \sin\left(\frac{k}{n^{2i/d}}\right) & \text{if } i \text{ is even} \\ \cos\left(\frac{k}{n^{2i/d}}\right) & \text{Otherwise} \end{cases}$$

- x: original embedding
- k: position in the time-series
- i: correspond to the $2i$ and $(2i + 1)$-th dimension of the embedding
- n: maximum lenth of input series
- d: embedding dimension, must be even

Positional Encoding

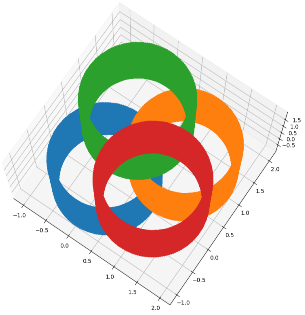

Figure 3: $n = 4096$

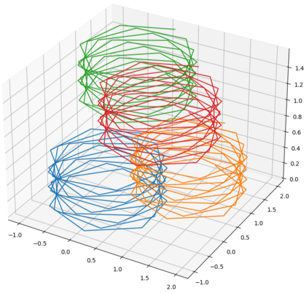

Figure 4: $n = 64$

Reinforcement Learning

- Pros
  - Aims to solve optimal control problems
  - Interactive environment allows us to take trading cost/price impact into account
  - More interpretable - the policy outputs distribution instead of single value
- Cons
  - Much greater variance
  - Requires more computation power (memory, cpu time, etc.) *We plan to train this in AWS Sagemaker.*

adaptive

Confidence Region

Problem

Find the Region in hedge ratio space that has a 95% chance including the best hedge ratio.

Note: the decision density π_t does not say anything about the best ratio, but definitely is a natural proxy for uncertainty

Conformal Prediction

- For each state S_t, we can estimate the best hedge $\hat{h}_t$ ratio by simulation.
- Define the score:

$$E_t := f(\pi_t, \pi_t(\hat{h}_t)) = \int_{\{a \in A | \pi_t(\hat{h}_t) <= \pi_t(a)\}} \pi_t(a)\, dA$$

- Simulate enough number of E_tś, and let

$$\hat{q} := \inf_{q \in \{E_t\}} \{ \frac{|\{E_t < q\}|}{|\{E_t\}|} > \alpha \}$$

be the α quantile of E_t ($\alpha = 0.95$)

adaptive

Conformal Prediction (Cont'd)

- For a new observation S_t and π_t corresponds to it, we may find

$$\{a \mid \tilde{p} <= \pi_t(a)\}$$

to be the desire confidence region, where

$$\tilde{p} = \sup_{p \in \mathbb{R}^+} \left\{ \hat{q} < \int_{\{a \in A \mid \pi_t(a) >= p\}} \pi_t(a) \, dA \right\}$$

- i.e. the inverse image $\pi_t^{-1}\left(f|_{\pi_t}^{-1}(\hat{q})\right)$ defines its boundary.

adaptive

Proximal Policy Optimization

- PPO is an state of art actor-critic framework that utilize policy gradient to optimize rewards according to the policy density and action advantage.
- Loss formulation:

$$L(s, a, \tilde{\theta}, \theta) = \min\left(\frac{\pi_\theta(a|s)}{\pi_{\tilde{\theta}}(a|s)} A^{\pi_{\tilde{\theta}}}(s, a), \underset{(1-\epsilon, 1+\epsilon)}{\text{clip}}\left(\frac{\pi_\theta(a|s)}{\pi_{\tilde{\theta}}(a|s)}\right) A^{\pi_{\tilde{\theta}}}(s, a)\right)$$

  - s - state
  - a - action
  - θ - policy parameter
  - $\tilde{\theta}$ - fixed policy parameter
  - A - advantage

adaptive

PPO Loss Visualization

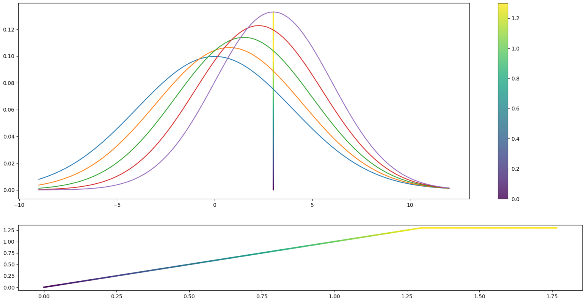

Figure 5: PPO Loss

Training Process

1. Data - historical price of options and underlyings, and other common data (e.g. risk free rate). Provided by *Algoseek*
2. Calibrate SDE models(e.g. Heston, Merton) parameters with historical data
3. Fit the parameter set to a joint distribution using Copulas
4. Iteratively train the neural networks:
    1. Simulate underlying dynamic and store/update the paths in agents memory pool
    2. Optimize the policy(actor) and baseline(critic) networks in an interleaving manner
        - Freeze the transformer encoder weights and fit the baseline network to discounted rewards
        - Unfreeze the transformer encoder weights and update policy network according to the PPO Loss
    3. Evaluate the out of sample performace

adaptive

Training Process (Cont'd)

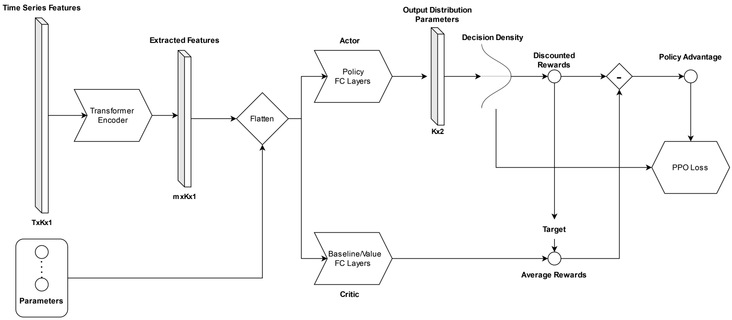

Figure 6: AI Hedging Architecture

adaptive

Validation Score

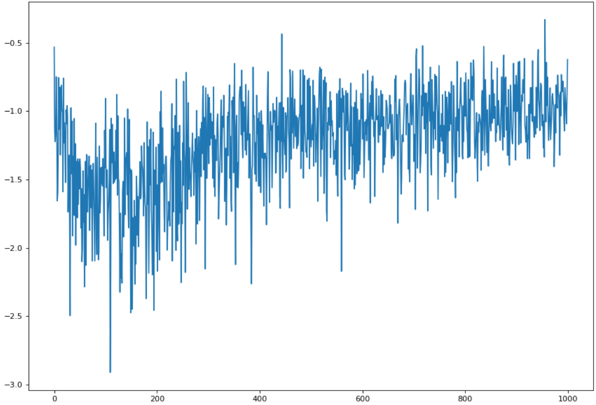

Figure 7: Average Downside

adaptive

Effect of Trading Cost

Top - Without trading cost
Bottom - With 1% commision cost

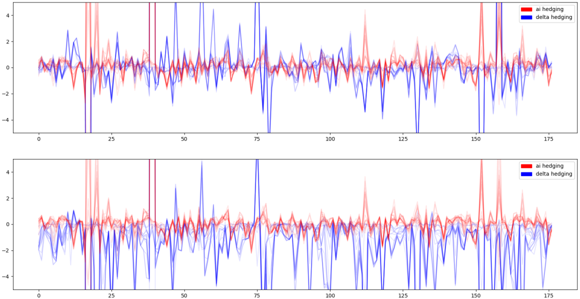

Figure 8: Wealth Process Comparison

Return Distributions

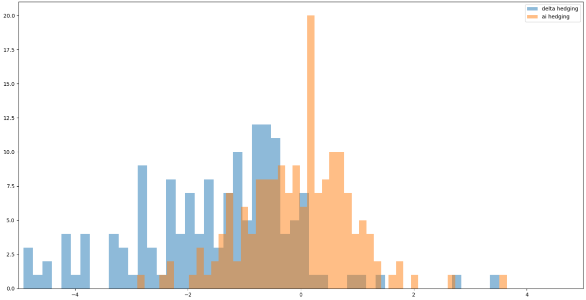

Figure 9: Distribution at Expiry

Return Distributions (Cont'd)

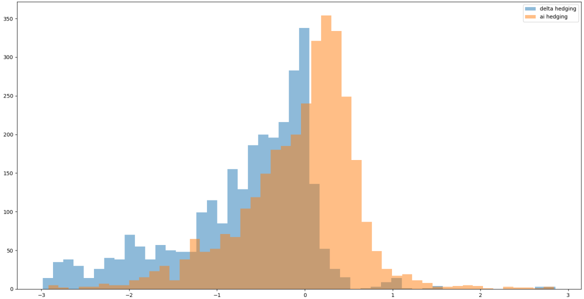

Figure 10: Full Path Distribution

Statistics

	Mean	Volatility	Skewness	Kurtosis	Quadratic HE
Delta Hedging	-0.25935	53.4690	10.33	226.50	10374.60
AI Hedging	0.46386	9.6676	24.65	750.51	5313.29

Down Side:

	Mean	Volatility	Kurtosis	CVaR	Quadratic HE
Delta Hedging	-3.53785	25.67	727.194	-40.8338	6038.9706
AI Hedging	-0.69568	1.8696	414.5817	-2.3849	95.8725

adaptive

Perfect Delta Does not Help!

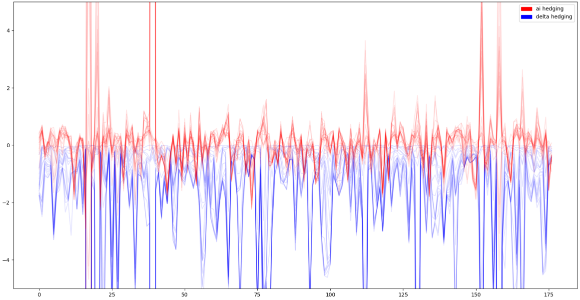

Figure 11: v.s. Perfect Delta

Future Applications

The Idea

Assume two hedgers, Alice and Bob, are hedging an derivative with underlyings. Alice takes a long position in option, while Bob takes the short position. The derivative price is fair if it makes both Alice and Bob undergoes same amount of residual risk.

A CVaR Example:

- Let the risk be measured using CVaR_α
- Denote the wealth process of two agents by $W_t^{\delta_a}(C_t, S_t)$, and $W_t^{\delta_b}(C_t, S_t)$, where δ_a and δ_b are strategies they follows, with ± 1 position fixed in the derivative C_t
- The Equal Risk Price C_0 is the value which satisfies the following equation:

$$\inf_{\delta_a}\{\mathrm{CVaR}_\alpha(W_T^{\delta_a}(C_t|C_0, S_t))\} = \inf_{\delta_b}\{\mathrm{CVaR}_\alpha(W_T^{\delta_b}(C_t|C_0, S_t))\}$$

adaptive

Risk Aversion Calibration

- The equal risk price depends on the level of risk aversion - α
- We could calibrate this parameter using historical data
- This could help filling in missing/arbitrage values in the option data
- However, this cannot be applied in real-time.
  - the optimization process is time consuming
  - One sided calibration is unstable

adaptive

Risk Metric and Pricing Kernel

- Utility and pricing kernel is another description of risk aversion
- The Euler equation of the consumption model implies

$$\mathrm{E}[\phi X] = \mathrm{Cov}[\phi, X] + \mathrm{E}[\phi]\,\mathrm{E}[X]$$

- Hence $\gamma := -r\,\mathrm{Cov}[\phi, X]$ is also a proxy for extent of risk aversion.
- Since both α and γ has a monotone relation with risk aversion, the bijection $\alpha \longleftrightarrow \gamma$ exists

adaptive

Risk Metric and Pricing Kernel (Cont'd)

Theorem (Ross Recovery)

If there is No Arbitrage, if the state price matrix P is irreducible, and if it is generated by a transition independent kernel, then there exists a unique (positive) solution to the problem of finding the natural probability transition matrix, F, the discount rate, δ, and the pricing kernel, ϕ. In other words, for any given set of state prices there is a unique compatible natural measure and a unique pricing kernel.

adaptive

Risk Metric and Pricing Kernel (Cont'd)

- We have

$$F = \left(\frac{1}{\delta}\right) DPD^{-1}, Fe = e$$

, Where $D = \frac{diag(\phi_{\theta_0 \to \theta_i})}{\delta}$

- Let $D^{-1}e = z$, we have $Pz = \delta z$
- We recognize this is the eigen-decomposition of P
- By Perron-Frobenius Theorem, P being a non-negative irreducible matrix, exists an unique positive eigen-value
- Since ϕ must be positive, it must be the eigen-vector corresponding to this largest eigen-value.

adaptive

Risk Metric and Pricing Kernel (Cont'd)

- Note the pricing kernel can be calculated without iterations
- If we may find any relation $\alpha \longleftrightarrow \gamma$ in the historical data, we could estimate the market's aggregated α value.
- This allow us to calculate risk equal price of derivatives in real time!

adaptive

References

Buehler, Hans and Lukas Gonon, Josef Teichmann, and Ben Wood (2018)
Deep Hedging
arXiv:1802.03042 [q-fin.CP]

Cao, Jay and Jacky Chen, John Hull, Zissis Poulos Joseph L. (2020)
Deep Hedging of Derivatives Using Reinforcement Learning
arXiv:2103.16409 [q-fin.CP]

Christoffersen and Peter and Heston, Steven L. and Jacobs, Kris (2009)
The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well
http://dx.doi.org/10.2139/ssrn.1447362

Cui, Yiran and Sebastian del Baño Rollin, Guido Germano (2016)
Full and fast calibration of the Heston stochastic volatility model
arXiv:1511.08718 [q-fin.CP]

Du, Jiayi and Muyang Jin, Petter N. Kolm, Gordon Ritter, Yixuan Wang, and Bofei Zhang (2020)
Deep Reinforcement Learning for Option Replication and Hedging
The Journal of Financial Data Science 5(3)

adaptive

References

Gammerman, A. and V. Vovk, and V. Vapnik. (1998)
Learning by transduction
In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence (UAI'98). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 148–155.

Heston, S. L. (1993)
A closed-form solution for options with stochastic volatility and applications to bond and currency options
The Review of Financial Studies, 6.2

Kazemnejad and Amirhossein, Inkit Padhi, Karthikeyan Natesan Ramamurthy, Payel Das, Siva Reddy (2023)
The Impact of Positional Encoding on Length Generalization in Transformers
arXiv:2305.19466 [cs.CL]

Kolm, Petter N. and Gordon Ritter (2018)
Dynamic Replication and Hedging: a Reinforcement Learning Approach
The Journal of Financial Data Science 1.1

Marzban, Saeed and Erick Delage, Jonathan Yumeng Li (2018)
Equal Risk Pricing and Hedging of Financial Derivatives with Convex Risk Measures
arXiv:2002.02876 [math.OC]

References

Peng, Xue Bin and Pieter Abbeel, Slevine, Michiel Van De Panne (2018)
DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills
arXiv:1804.02717 [cs.GR]

Ross, Stephen A. (2005)
Neoclassical Finance
Princeton Univeristy Press

Ross, Stephen A. (2015)
The Recovery Theorem
The Journal of Finance, 70.2

Ruf, Johannes and Weiguan Wang (2021)
Hedging with Linear Regressions and Neural Networks
Forthcoming in *Journal of Business and Economic Statistics*

Schulman, John and Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimovs (2017)
Proximal Policy Optimization Algorithms
arXiv:1707.06347 [cs.LG]

Vaswani, Ashish and Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin (2020)
Attention Is All You Need
arXiv:1706.03762 [cs.CL]

adaptive

THANK YOU

Adaptive Investment Solutions, LLC
6 Liberty Square, No. 2352, Boston MA 02109
www.adaptive-investments.com
hello@adaptive-investments.com
+1.781.728.9556 • +1.212.337.9714