Dynamic Liquidation Strategy for Large Quantity of Stock using Reinforcement Learning

Hwang, Jo, Lee, Mok

Global Economics, Sungkyunkwan University

January 8, 2024

Outline

Overview of Liquidation Problem

- Suppose we want to sell a large quantity of a specific stock in one day.

- We have two naive liquidation strategies.

  1. First Strategy: Sell everything once
  2. Second Strategy: Sell small amount many times

- Two Strategies have trade-off

  1. Market Impact would be huge, but no unpredictable risk
  2. Market Impact would be small, but exposed to unpredictable risk

- We need to *optimize this trade-off* and decide our trading strategy.

Overview of Liquidation Problem

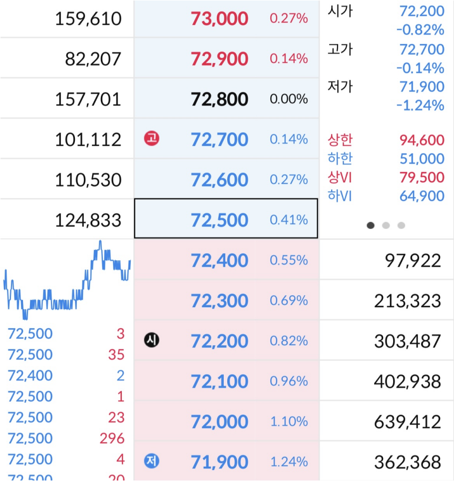


Figure: Example of Order Book

Overview of Liquidation Problem

- Trading firms generally use TWAP and VWAP strategies.
  1. TWAP strategy: Executing trades evenly over a specified time period.
  2. VWAP strategy: Executing more orders as volume increases
- However, these two static strategies determine their trading strategies before trading starts.
- They do not consider the *price change* in their trading period.

Research Question

- We want to make a *Dynamic Liquidation Strategy* which considers price change in given trading period.
- We used *Reinforcement Learning* method to apply price change into our trading strategy.
- We adopted methodology of 'Amgren-Chriss Model' for trading strategy which is a fundamental paper that has been quoted over 2,000 times.

Why Reinforcement Learning?

- Our Liquidation Problem is such a "Sequential Decision Problem".
- This process is similar with how the reinforcement learning learns

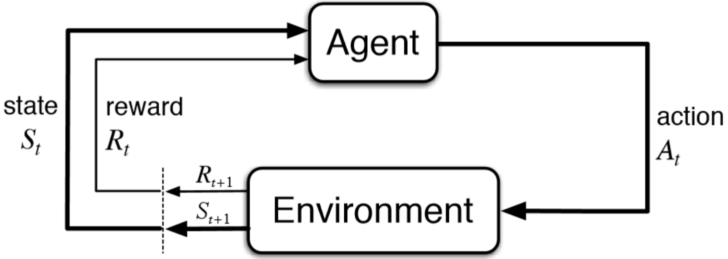


Figure: Framework of reinforcement learning

- **Goal : Maximize** $G_t = R_{t+1} + \gamma R_{t+2} + \ldots + \gamma^{T-t} R_{T+1}$
- The policy of this agent will be our trading strategy

Almgren and Chriss Model (1)

Price Dynamics and Market Impact

- Permanent Market Impact: Let S_k be the price at time t_k

$$S_k = S_{k-1} + \sigma\tau^{1/2}\xi_k - \tau g(v)$$

- Temporary Market Impact: $\tilde{S}_k$ is an actual execution price

$$\tilde{S}_k = S_{k-1} - h(v)$$

Cost of Trading Strategy

- Let X be total share and n_k be the quantity of stock we sold for $\tilde{S}_k$
- Total cost of trading is

$$\mathcal{C} = XS_0 - \sum_{k=1}^{N} n_k \tilde{S}_k = \sum_{k=1}^{N} x_k \left(\sigma\sqrt{\tau}\xi_k - \tau g_k\right) - \sum_{k=1}^{N} n_k h_k$$

Almgren and Chriss Model (2)

Mean Variance Analysis

- Take expectation and variance for cost and make utility function

$$U(\mathcal{C}) = \mathbb{E}[\mathcal{C}] + \lambda \operatorname{Var}[\mathcal{C}]$$

- λ is risk aversion parameter of trader

Optimal Solution

- They get static optimal trading strategy n_k by using FOC $\frac{\partial U}{\partial n_k} = 0$

$$n_k = \frac{2 \sinh\left(\frac{1}{2}\kappa\tau\right)}{\sinh(\kappa T)} \cosh\left(\kappa\left(T - t_{k-\frac{1}{2}}\right)\right) X \text{ for } k = 0, \ldots, N$$

- We will use their utility function and static trading strategy to find dynamic optimal trading strategy

Modeling Reinforcement Learning - States

- Our state vector must contain some information about the time remaining, or what is equivalent, the number trades remaining.

$$[r_{k-5}, r_{k-4}, r_{k-3}, r_{k-2}, r_{k-1}, r_k, m_k, i_k]$$

where

- $r_k = \log\left(\frac{\tilde{S}_k}{\tilde{S}_{k-1}}\right)$ is the log-return at time t_k
- $m_k = \frac{N_k}{N}$ is the number of trades remaining at time t_k normalized by the total number of trades.
- $i_k = \frac{x_k}{X}$ is the remaining number of shares at time t_k normalized by the total number of shares.

Modeling Reinforcement Learning - Actions

- We set action a_k as a percentage of the solution of AC suggested
- By giving near-opimal solution, we can reduce traning time
- We set a_k produced by agent between 0.5 and 1.5.
- Using this method, agent determine n_k, *the number of shares* to sell at each time by

$$n_k = a_k \times AC_k$$

- where AC_k is the number of stocks to sell at time t_k at AC paper

Modeling Reinforcement Learning - Rewards

- Since we don't know which order is good exactly, we give agent a clue for which order is better.
- We give more reward to agent when the utility (of agent's remaining stocks) decreases
- x_t^* is remaining stocks at time t

$$R_t = \frac{U_t\left(x_t^*\right) - U_{t+1}\left(x_{t+1}^*\right)}{U_t\left(x_t^*\right)}$$

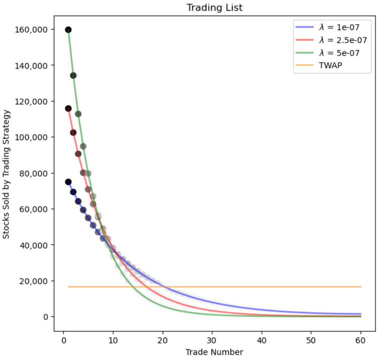
- Since more reward is given as $U(C)$ decreases, agent will action toward minimizing market impact

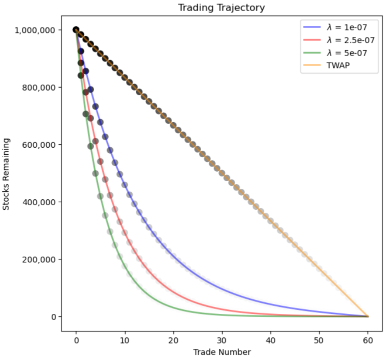
Simulation

- Total shares we need to sell : 1 million
- Initial Price : \$50
- Price Dynamics : Arithmetic discrete random walk
- Number of available trade : 60
- Daily Volatility : 0.8
- Fixed cost of selling per share : \$0.062
- Risk aversion parameter : 10^{-7}, 2.5×10^{-7} and 5×10^{-7}

Result (1)

- Our agent learns 10000 episodes in this simulation environment





- Half Life of our strategy is shorter than TWAP
- We can guess risk aversion of TWAP is lower than others

Result (2)

- We get mean and standard deviation of trading cost for each models and TWAP by simulating 10000 times

Trading Cost	**Mean**	**STD**
TWAP	210,220	1,661,784
$\lambda = 10^{-7}$	268,004 (+27.5%)	912,361 (-45%)
$\lambda = 2.5 \times 10^{-7}$	324,881 (+54.5%)	707,578 (-57.4%)
$\lambda = 5.0 \times 10^{-7}$	387,967 (+84.5%)	578,764 (-65.1%)

Table: Result of our model and TWAP

- Trading Cost of our models are higher than TWAP but the risks lower
- As the mean goes down, the variance goes up

Analysis (1)

- We can capture the trade-off between two naive strategies.

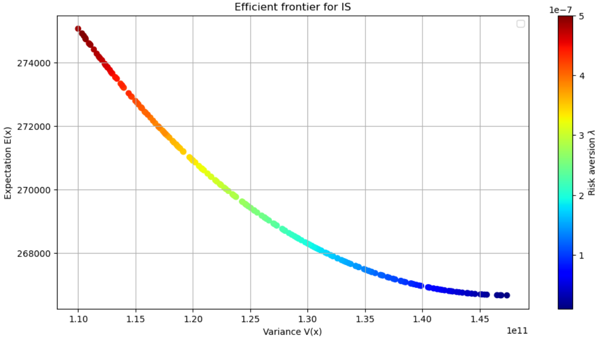


Figure: Efficient Frontier of our simulation

- As variance goes higher, expectation of cost goes lower
- As risk aversion parameter(λ) goes lower, variance goes higher

Analysis (2)

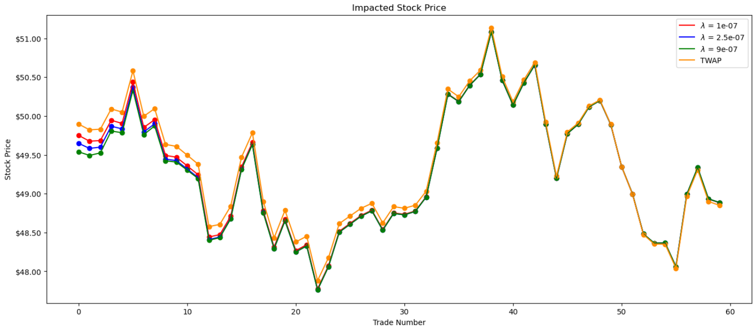


Figure: Simulated price of stock affected by each strategy's market impact

- Market Impact of TWAP is lower than other model
- But since they should sell same amount even later period, so variance increases

Conclusion and Limitation

- We confirmed trade-off exists between expactation and variance of trading cost
- Our dynamic model is less risky than TWAP.
- Therefore, for some traders whose risk aversion is high, our dynamic liquidation strategy can be an alternative to TWAP strategy
- Also, trader can customize our dynamic strategy by easily adjusting risk aversion parameters.
- But our model cannot be used as an alternative for VWAP since they do not consider volume at all.
- We need voluminous real world data to simulate volume like price. (ex. minute data)

Reference

- Robert Almgren and Neil Chriss, Optimal execution of portfolio transactions, Journal of Risk 3 5–40 (2001)
- Jim Gatheral and Alexander Schied, Optimal Trade Execution under Geometric Brownian Motion in the Almgren and Chriss Framework, International Journal of Theoretical and Applied Finance 14(3) 353–368 (2011)
- P.A. Forsyth, J.S. Kennedy, S. T. Tse, and H. Windcliff, Optimal Trade Execution: A Mean- Quadratic-Variation Approach, University of Waterloo (2011). Jim Gatheral, No-dynamic-arbitrage and market impact, Quantitative Finance 10(7) 749–759 (2010).
- D. Hendricks and D. Wilcox, "A reinforcement learning extension to the Almgren-Chriss framework for optimal trade execution," 2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr), London, UK, 2014, pp. 457-464, doi: 10.1109/CIFEr.2014.6924109.