Do We Really Need Reinforcement Learning in Algorithmic Trading?

Davide Maran

What is algorithmic trading?

Financial assets

- **Stocks (Apple, Nvidia, Microsoft)**
- Currencies (EUR/USD)
- Commodities (oil, gas, gold)
- Bonds

What is algorithmic trading?

Financial assets

- **Stocks (Apple, Nvidia, Microsoft)**
- Currencies (EUR/USD)
- Commodities (oil, gas, gold)
- Bonds

An asset is characterized by a price p_t.

What is algorithmic trading?

Financial assets

- **Stocks (Apple, Nvidia, Microsoft)**
- Currencies (EUR/USD)
- Commodities (oil, gas, gold)
- Bonds

An asset is characterized by a price p_t.

Algorithmic trading

Using software to decide when to buy or sell.

A Naïve view of Algorithmic Trading

$h-$step return:

$$\Delta_h p_t := p_{t+h} - p_t.$$

A Naïve view of Algorithmic Trading

$h-$step return:

$$\Delta_h p_t := p_{t+h} - p_t.$$

Simple algorithmic trading pipeline

- Collect historical data
- Train regression algorithm $\widehat{p}_{t+h}$
- Buy if $\widehat{p}_{t+h} > p_t$, sell otherwise.

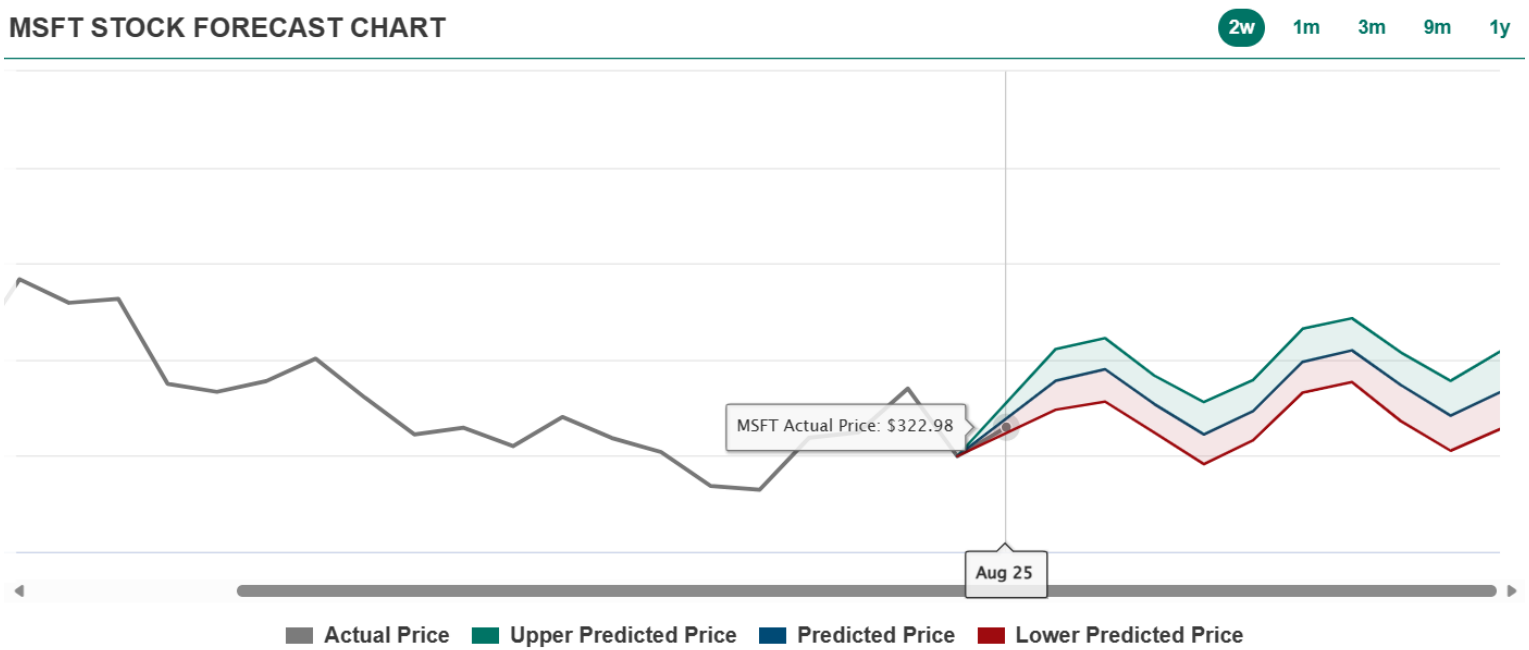
A Naïve view of Algorithmic Trading

$h-$step return:

$$\Delta_h p_t := p_{t+h} - p_t.$$

Simple algorithmic trading pipeline

- Collect historical data
- Train regression algorithm $\widehat{p}_{t+h}$
- Buy if $\widehat{p}_{t+h} > p_t$, sell otherwise.



... But the picture is deeply incomplete

Idea	**Reality**
"Buy at p_t"	Impossible to buy at midprice

... But the picture is deeply incomplete

Idea	**Reality**
"Buy at p_t"	Impossible to buy at midprice
"Maximize expected return"	Position limits, risk constraints

... But the picture is deeply incomplete

Idea	**Reality**
"Buy at p_t"	Impossible to buy at midprice
"Maximize expected return"	Position limits, risk constraints
"Return $= p_{t+h} - p_t$"	Latency, transaction costs and market impact

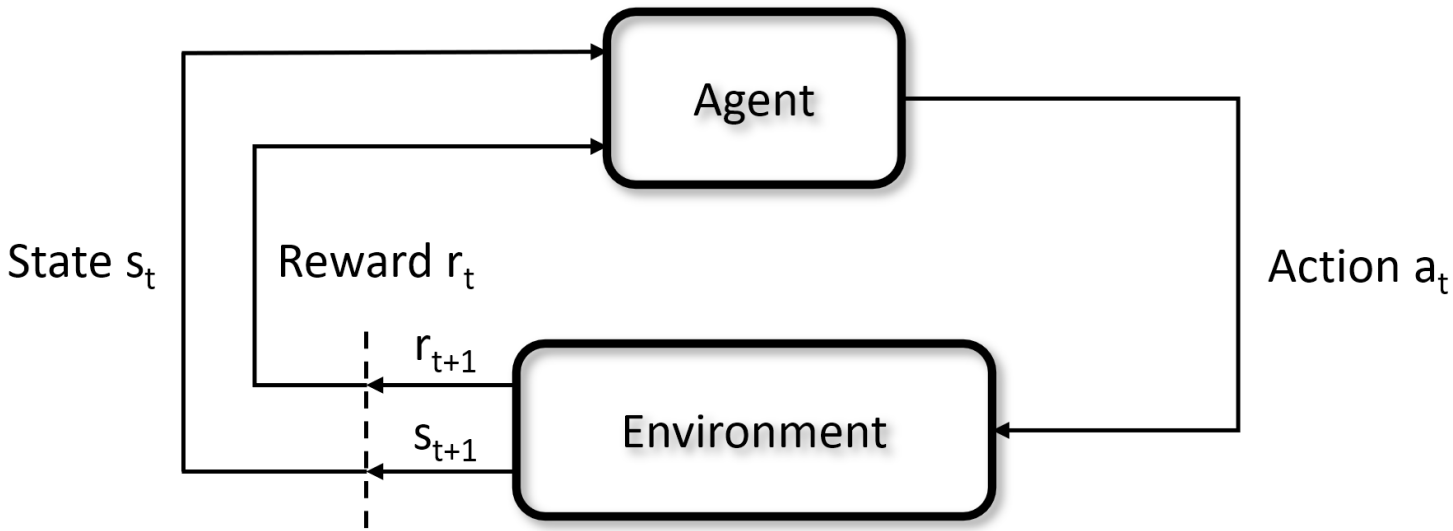
... But the picture is deeply incomplete

Idea	**Reality**
"Buy at p_t"	Impossible to buy at midprice
"Maximize expected return"	Position limits, risk constraints
"Return $= p_{t+h} - p_t$"	Latency, transaction costs and market impact
"Better training accuracy $\Longrightarrow$ Better performance"	Market is non-stationary

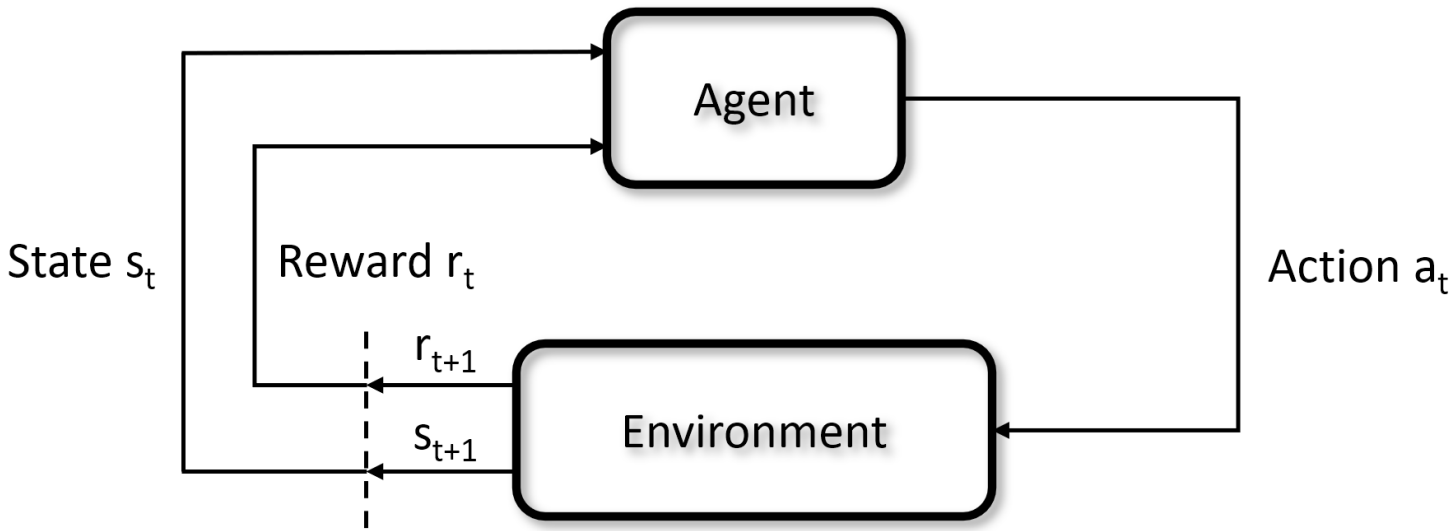
Trading is not just prediction

Do we need sequential decision-making methods such as Reinforcement Learning?

Reinforcement Learning in a nutshell

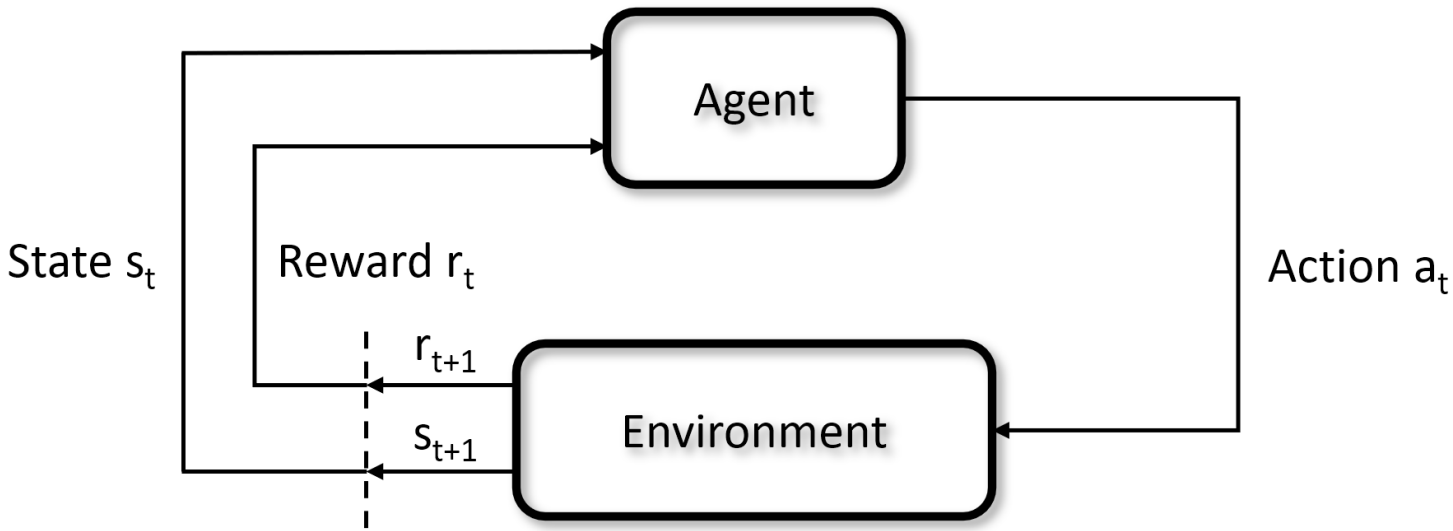


Reinforcement Learning in a nutshell



1. See s_t
2. Play a_t
3. State evolves to s_{t+1} and reward r_{t+1} is collected

Reinforcement Learning in a nutshell



1. See s_t
2. Play a_t
3. State evolves to s_{t+1} and reward r_{t+1} is collected

Goal: maximize

$$\sum_{t=1}^{\infty} \gamma^t r_t \qquad \gamma \in (0,1).$$

Reinforcement Learning Milestones

Position and Risk constraints

Typical scenario

The rules of your hedge fund impose

- Position limits: no more than 10M\$ and not less than -5M\$
- Daily loss limit: -1M\$
- Neutral position at the end of the day

Position and Risk constraints

Typical scenario

The rules of your hedge fund impose

- Position limits: no more than 10M$ and not less than -5M$
- Daily loss limit: -1M$
- Neutral position at the end of the day

Can RL deal with it?

Position and Risk constraints

Typical scenario

The rules of your hedge fund impose

- Position limits: no more than 10M$ and not less than -5M$
- Daily loss limit: -1M$
- Neutral position at the end of the day

Can RL deal with it?

Easy ✓

Include position + current loss in the environment state; position limits in the reward function.

Spread and Execution Costs

Agent cannot buy at the midprice

Market price = midpoint between the best buy and sell orders.

- To buy: look at sell orders
- To sell: look at buy orders

Spread = lowest "sell" - highest "buy".

Spread and Execution Costs

Agent cannot buy at the midprice

Market price = midpoint between the best buy and sell orders.

- To buy: look at sell orders
- To sell: look at buy orders

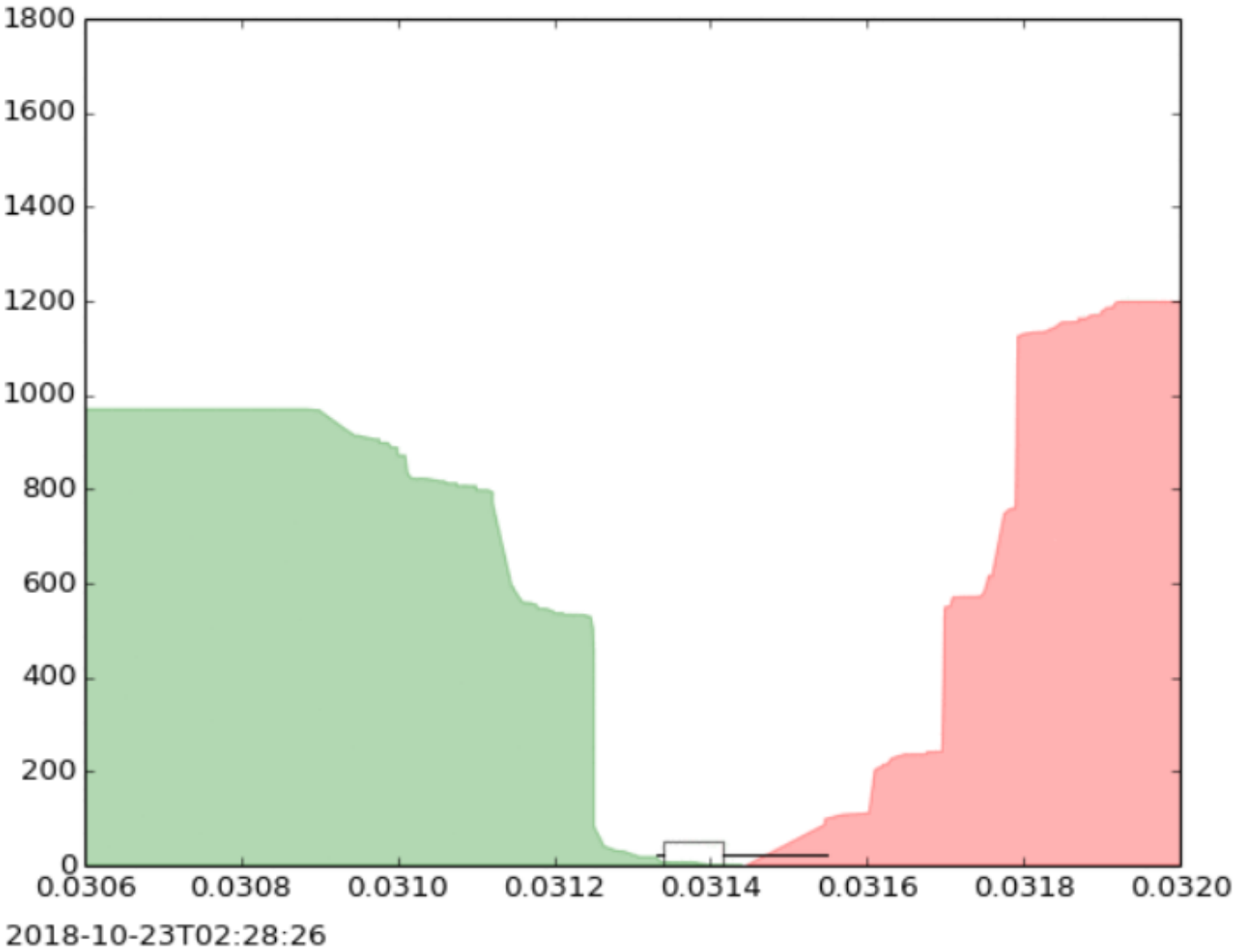
Spread = lowest "sell" - highest "buy".

Latency costs

Between decision time and execution time:

- prices may change,
- competing agents may react,
- liquidity may disappear.

Order book



- Red = sell orders, Green = buy orders
- Price on x-axis, Amount on y-axis

Spread and Execution Costs

Agent cannot buy at the midprice

[...]

Latency costs

[...]

Can RL deal with it?

Spread and Execution Costs

Agent cannot buy at the midprice

[...]

Latency costs

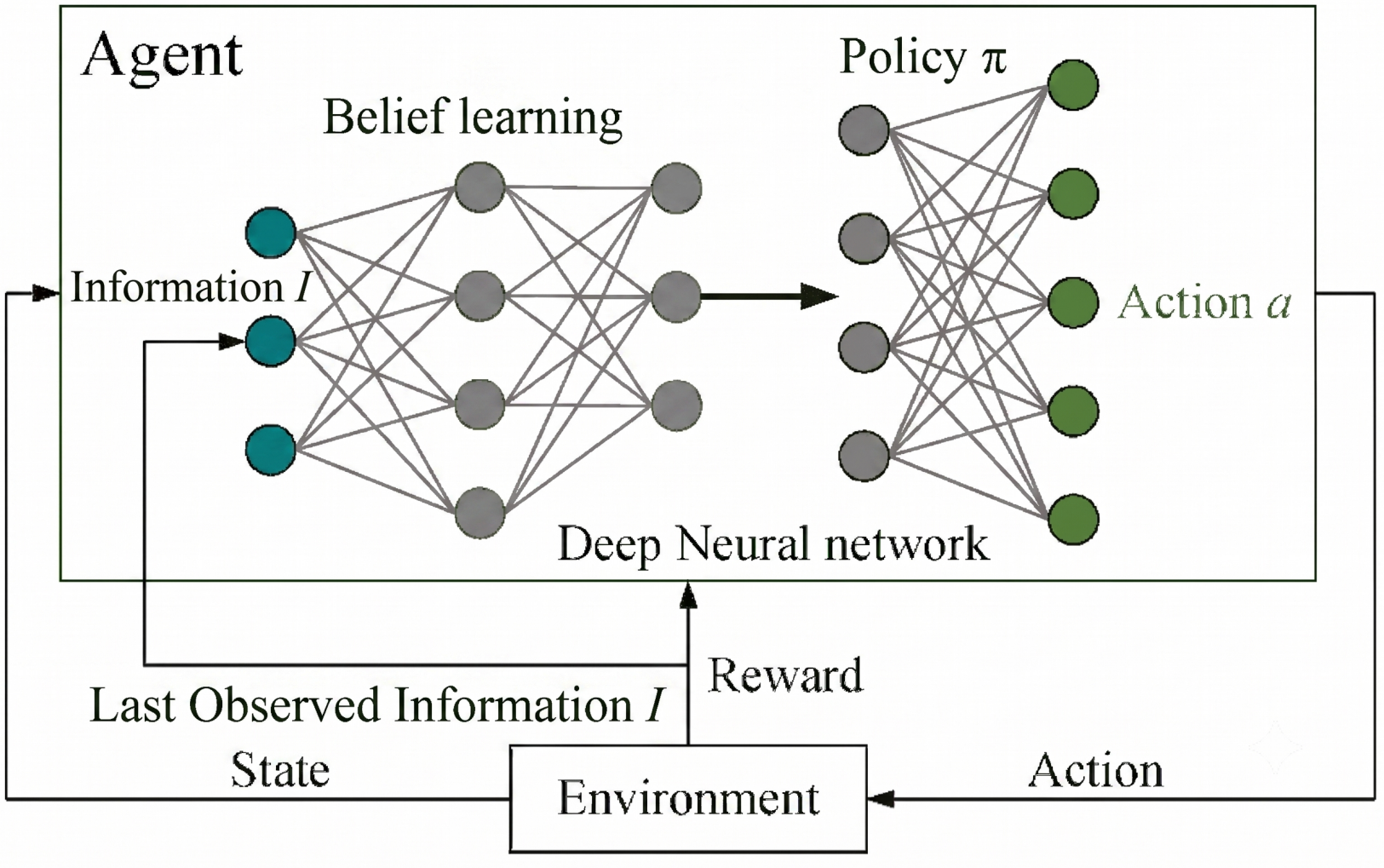
[...]

Can RL deal with it?

Well-studied problem ✓ (Delayed RL)

Add spread to the state and use delayed RL techniques [Liotet, 2023, Schuitema et al., 2010, Firoiu et al., 2018, Bouteiller et al., 2020, Liotet et al., 2022, Wang et al., 2024].

Policy Learning under Delay



Market Impact

Two types of impact [Almgren and Chriss, 2001]

When buying many stocks ($10M$\$ for AAPL, $100k$\$ for small companies)

Market Impact

Two types of impact [Almgren and Chriss, 2001]

When buying many stocks ($10M\$$ for AAPL, $100k\$$ for small companies)

- *Transient impact*: average price is higher than when buying one unit.

Market Impact

Two types of impact [Almgren and Chriss, 2001]

When buying many stocks ($10M\$$ for AAPL, $100k\$$ for small companies)

- *Transient impact*: average price is higher than when buying one unit.
- *Permanent impact*: initially the price is pushed up, then MM react adversarially.

Hard problem ✓/✗

Transient impact $\in$ stochastic reward.

Market Impact

Two types of impact [Almgren and Chriss, 2001]

When buying many stocks ($10M\$$ for AAPL, $100k\$$ for small companies)

- *Transient impact*: average price is higher than when buying one unit.
- *Permanent impact*: initially the price is pushed up, then MM react adversarially.

Hard problem ✓/✗

Transient impact $\in$ stochastic reward. Permanent impact requires multi-agent + partial observability $\Longrightarrow$ difficult.

Non-stationary market

Non-stationairty

Patterns observed in the past often do not repeat.

- Many agent use models trained on the **same** past data.

Non-stationary market

Non-stationairty

Patterns observed in the past often do not repeat.

- Many agent use models trained on the **same** past data.

Hard ✗

- Impossible to do exploration efficiently.
- Loop feedback complicates everything.

Is RL the solution to algo trading?

Why ✓

- Position and Risk constraints
- Execution costs
- Transient market impact

Is RL the solution to algo trading?

Why ✓

- Position and Risk constraints
- Execution costs
- Transient market impact

Why ✗

- Permanent impact
- Non-stationary market

Is RL the solution to algo trading?

Why ✓

- Position and Risk constraints
- Execution costs
- Transient market impact

Why ✗

- Permanent impact
- Non-stationary market

RL cannot do all the work

May work in combination with a model for the market impact and robust price predictors.

Bibliography

Almgren, R. and Chriss, N. (2001).
Optimal execution of portfolio transactions.
Journal of Risk, 3:5–40.

Bouteiller, Y., Ramstedt, S., Beltrame, G., Pal, C., and Binas, J. (2020).
Reinforcement learning with random delays.
In *International conference on learning representations*.

Firoiu, V., Ju, T., and Tenenbaum, J. (2018).
At human speed: Deep reinforcement learning with action delay.
arXiv preprint arXiv:1810.07286.

Liotet, P. (2023).
Delays in reinforcement learning.
arXiv preprint arXiv:2309.11096.

Liotet, P., Maran, D., Bisi, L., and Restelli, M. (2022).
Delayed reinforcement learning by imitation.
In *International conference on machine learning*, pages 13528–13556. PMLR.

Schuitema, E., Buşoniu, L., Babuška, R., and Jonker, P. (2010).
Control delay in reinforcement learning for real-time dynamic systems: A memoryless approach.
In *2010 IEEE/RSJ international conference on intelligent robots and systems*, pages 3226–3231. IEEE.

Wang, W., Han, D., Luo, X., and Li, D. (2024).
Addressing signal delay in deep reinforcement learning.
In *International Conference on Learning Representations*, volume 2024, pages 15873–15897.